

Dealing with democracy

The drive for greater public participation in the regulation and politics of technologies is both necessary and irreversible. But proposals to extend it into the selection of publicly funded research contain dangers to science and society.

The increasing involvement of the public in the policy-making and regulation of science and technology is to be welcomed. As a recent forum of supportive discussions describes (see *Science and Public Policy* **30**, 146–225; 2003), governments, parliaments and other public organizations in France, Germany and Britain are taking significant steps in this direction, no doubt partly inspired by models such as consensus conferences pioneered in Scandinavia. The European Commission has published a White Paper extolling the virtues and processes of public participation, and is implementing it in its Sixth Framework research programme.

As the forum also makes clear, this is in increasingly sharp contrast to the United States where, despite enviable traditions of stakeholder involvement, the Bush administration is moving against public access to decision-making and towards the biasing of advisory bodies to suit its own agendas. Advocates of public participation, such as the Loka Institute, find themselves further out in the wilderness than they have been for a long while.

The European shift towards broad public participation is an appropriate move against the idea that experts always know best. That idea underpins the 'deficit model' — the label given by professionals in the business of public understanding and consultation to the mistaken belief that the public's mistrust of science and technology can be removed simply by explaining scientific concepts, processes and facts.

These same professionals are all too quick to regurgitate the cliché that the deficit model is dead. Indeed it is, and few scientists who have any experience of societal distrust would nowadays maintain that scientific reassurance and knowledge are all that are needed to remove it. But even so, crippling deficits of knowledge and understanding remain, between scientific disciplines and, in both directions, between science and its stakeholders. The goal of increased public accountability and participation serves only to highlight such deficits. This represents a major but unavoidable challenge for governments, who are ultimately responsible for ensuring that citizens can obtain timely and comprehensible information about politically controversial science and technology from trustworthy sources.

More involvement

Expertise is not confined to researchers, however. Whatever their stance, lobbyists, patients' groups and environmental stakeholders often take the time to explore and question the literature in depth. And there is no inevitable polarization between researchers and the public. Rather, polarization can reflect shared attitudes towards technologies such as embryonic stem cells, reproductive cloning and nuclear fission reactors, or towards burdens of proof about the impact of technologies in decisions about regulation.

Despite the evident progress in formalized public participation and discussion, there are those who say that this is not enough, and that the selection of science for funding should also be subject to public involvement. In their contribution to the forum, policy researchers Reiner Grundmann and Nico Stehr propose that science-related politics should be reinvented from the ground up. This, they say, is because the power of science to alter nature has reached such a state that society needs to have a much more fundamental place in

considering its support. Sue Mayer, director of the lobby group Genewatch UK, takes a similar line. Inadequate representation of the wider public in the funding and intellectual-property systems, she says, leaves them with insufficient scope to "influence the trajectory or accessibility of scientific information".

Many will perceive proposals for further democratic involvement as an anti-science Trojan horse standing outside science's citadel. The proposals would certainly need critical scrutiny, as they bear all the hallmarks of advocacy rather than analysis. They understate the existing means by which scientific communities are subject to democratic discussions and regulation, and oversimplify episodes that they invoke as symptoms of scientists' resistance to restraint. They ignore the subtleties of the way in which fundamental science is unpredictable, unavoidably sets its own agendas, and has an inherent timescale, both in its community structure and its execution, that is ill matched to the short-term perceptions of public opinion. With such shortcomings, they carry dangers both for science and for society.

But the proposals have virtues too. There is a strong case for funding agencies to include people who are particularly motivated to encourage research into controversial science and its applications on a broader basis than the pursuit of national wealth or power, or 'sweet technology' for its own sake.

A balanced approach

In short, science's structures need to avoid procedural savagery equivalent to the destruction of trials of genetically modified (GM) crops and physical attacks on animal researchers, while taking democratic steps that, by increasing society's trust in the pursuit of scientific insights, lessen the likelihood of such vandalism. This week sees the publication of the results of an unprecedented public consultation about GM crops, instigated by the UK government (see page 331). There will no doubt be headlines about the British public's well-known scepticism of the technology. But attention also needs to be paid to the strongly held view emerging from the survey that more research on GM crops needs to be done.

Proposals for more democracy also deserve critical scrutiny with regard to the process. There has been much public participation over the years in many countries, and remarkably little analysis of its effectiveness and participants' experiences. For example, too many citizen's juries seem to align themselves with the positions of those who commissioned them. Glib advocacy for more and deeper 'democratization' is no longer good enough. Critical evaluation is needed of the successes and dangers in the process.

Are we all experts now? Is this the century of the informed, competent, and ever more emancipated global expert-citizen? To the extent that, thanks to the Internet and public consultations, the answers to these questions are increasingly affirmative, the scientific community needs to adapt to the public's need for accountability and trust. To the extent that the answers to both are still negative, science needs to defend itself against threats from fundamentalism and ignorance. This balance between adaptation and resistance will challenge scientific communities for the foreseeable future. But both will be required. ■





Sharp rebuke
Biodefence scientists
make disputed claims
for smallpox vaccine
p332



High hopes
New Chinese agency
aims to transform
biomedical research
p333



Access denied
Paper ignites row
with journal over
open access
p334



Sail on
Ocean drilling project
relaunched on
international waters
p337

Gates steps up war on malaria with donation of \$168 million

Declan Butler, Manhica, Mozambique

Bill and Melinda Gates, the world's richest couple, are to make the biggest-ever single donation to malaria research.

The Bill and Melinda Gates Foundation, which has an endowment of \$25 billion, is to provide a \$168-million block of grants for various treatment and vaccination strategies. It has already donated more than \$3 billion to global health research.

The philanthropists made the announcement on a visit to Mozambique last weekend, at the Manhica Health Research Centre 80 kilometres north of the capital, Maputo, accompanied by the prime minister Pascoal Mocumbi. Earlier, the couple held children — some of them suffering from cerebral malaria — taking part in the largest clinical trial of a malaria vaccine.

Bill Gates said that their trip to Africa, which included visits to South Africa and Botswana to discuss AIDs-related issues, was to raise awareness of the need for more research into malaria. The disease infects almost half a billion people and kills more than a million annually, mostly young children in Africa. The resurgence of malaria in Africa is partly a result of increased resistance to common antimalarial drugs.

Current spending on malaria control is \$200 million a year, far short of the \$2.5 billion a year that the WHO Commission on Macroeconomics and Health estimates is needed.



Safe hands: Bill Gates with baby Cecil Massango.

Total annual research spending on malaria is just \$100 million. Not everything will be solved by his money, says Gates, adding that his emphasis is to bridge the gap between basic research and industrial development.

Of the new grants, \$28 million over five years will fund an international consortium to test as quickly as possible a promising new control technique known as 'intermittent

preventive treatment'. Two years ago researchers reported that incidence of severe malaria was halved in infants who were given an antimalarial drug three times during their first year of life, when they attend routine immunization sessions (D. Schellenberg *et al. Lancet* **357**, 1471–1477; 2001). The natural immunity built up by infected children was not impaired by the treatment.

Richard Klausner, head of the foundation's global health activities, says that the observations are scientifically fascinating. "The children continue to show the memory of being given treatment," he says, as if the drug were almost working as a vaccine. Fifteen studies in nine countries will now explore this further, with a view to rapid implementation if it works.

The public-private drug-development partnership Medicines for Malaria Venture, of Geneva, will receive \$40 million to accelerate development of the four most advanced drugs in its 21-candidate pipeline. Its goal is for at least one new antimalarial drug to be licensed by the end of the decade.

The Seattle-based Malaria Vaccine Initiative is awarded \$100 million over four years to speed clinical trials of its 15 candidates.

The foundation is discussing grants for a malaria postgenomics programme with the US National Institutes of Health and the UK Wellcome Trust, Klausner adds, and envisages grants for mosquito-vector research. ■

UK public opposes government on transgenic crops

Jim Giles, London

The British government, set to rule on the legalization of genetically modified (GM) crops, is in a corner after a study that it commissioned confirmed deep-seated concern among the public.

Although the government is believed to want to allow transgenic crops to be grown in Britain, it will find the negative report hard to ignore, in part because of recent accusations that it ignored public opinion over its decision to go to war with Iraq.

The results of this summer's 'GM Nation?' debate, a series of some 670 public meetings that attracted up to 20,000 participants, reveal a British public uneasy with the potential environmental, health and socio-economic effects of transgenic crops. More than half the participants who returned feedback forms after the debate said they never wanted to see GM crops grown in Britain, according to a report due to be released on 24 September.

The government commissioned the

exercise to help it decide whether to give the go-ahead for commercialization of transgenic crops. Britain, along with other European Union countries, will begin voting later this year on whether to license some 15 new varieties of transgenic crop, ending a five-year moratorium on new approvals.

Industry representatives say that the results of the debate were skewed by the presence of environmental activists at the meetings, but the result may still put pressure on the government. ■

Report raises hopes for grand network of US ecology centres

Rex Dalton, San Diego

The US National Academy of Sciences (NAS) has endorsed the creation of an ambitious network of ecological research stations that the National Science Foundation has been advocating for the past six years.

The National Ecological Observatory Network (NEON) will provide a system for continental-scale experiments on ecological systems and will assist in predicting environmental change.

Congressional funding for NEON — whose observatories could cost \$20 million each to build, plus \$3 million in annual operating costs — has repeatedly been stalled for political or budgetary reasons. Supporters hope that the NAS report, completed last week after four months of study, will encourage Congress to fund the \$12-million first phase during the 2004 federal fiscal year beginning on 1 October. A funding decision will be made within the next few weeks.

The 77-page NAS report, compiled by a panel of 14 scientists, recommends that NEON should concentrate on six major ecological observatories: biodiversity, biogeochemical cycles, climate change, infectious diseases, invasive species and habitat alternation.

The NSF had initially proposed 17 observatories, but the panel says that concentrating on six ecological issues would allow a better focus. "We think it will produce better science this way," says David Tilman, a plant ecologist at the University of Minnesota in Minneapolis-St Paul, who chaired the NAS panel.

A NEON observatory would include a core site linked by the Internet to satellite facilities and complementary institutions such as a university, government lab or museum. The core site would concentrate instrumentation for data collection, processing and analysis. Scientists in the complementary organizations could then study a target ecological issue in a variety of environments, from grasslands to forests and deserts to waterways.

John Aber, a biogeochemist at the University of New Hampshire in Durham and a NAS panel member, says that the committee has given the NSF flexibility to create NEONs through competition.

"The first NEON observatory should be national in scope," says Aber. But it will be up to the NSF to decide whether competitions should be launched in all six areas, or in only one, in the first round of funding, he adds. ■



Lab jab: researchers have claimed in a press release that smallpox vaccination boosts HIV immunity.

Vaccine claim lifts company's stock but angers researchers

Erika Check, Washington

A press-released claim that the smallpox vaccine could increase immunity to HIV has boosted share prices of the vaccine's manufacturers — but has been met with hostile scepticism by the scientific community.

The claim, made by researchers at George Mason University in Fairfax, Virginia, on 11 September, has not been substantiated by published data. Other scientists are particularly annoyed at the unconventional method of the announcement, because of the claim's potential implications for past and future smallpox-vaccination programmes.

Raymond Weinstein, a researcher at George Mason's National Center for Bio-defense, who led the research, says he noticed that the countries where smallpox had been eradicated first now have the highest rates of HIV infection. Together with Ken Alibek, a former Soviet bioweapons scientist who is now at George Mason, he directed a study by microbiologists at George Washington University Medical Center in Washington DC. Blood was tested from ten volunteers who had been vaccinated against smallpox, and ten who had not. The team claims that blood cells from vaccinated subjects became infected with four times less HIV than cells from unvaccinated subjects. Smallpox and HIV can both use the same cell-surface protein to infect cells, so it is plausible that there may be a link between immunity to the two diseases, claims Weinstein.

Weinstein says that the team then briefed

officials at the National Institutes of Health and its parent agency, the Department of Health and Human Services. They decided to announce their findings in a press release, he says, because federal scientists warned them that they might try to race a study of their own into publication.

But Ed Tramont, director of the Division of AIDS at the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, says that he is not aware of any scientists working on duplicating the results. And he cautions against overinterpreting the press-released data. Many vaccinations cause temporary boosts in immunity soon after patients receive them, he says, and many adults who received the smallpox vaccine as children have subsequently contracted HIV.

Donald Henderson, a major architect of the smallpox-eradication plan and an adviser to the federal government on bioterrorism preparedness, says that there is in any case no consistent relationship between smallpox eradication and HIV emergence in Africa. "It doesn't make a lot of sense," he says.

But on the day that Weinstein and Alibek made their announcement, shares in smallpox-vaccine manufacturer Acambis jumped by 10% to a new high. And even though vaccine-maker VaxGen had announced earlier this year that its experimental HIV vaccine failed in clinical trials, its shares also rose by 10%. ■

SARS triggers biomedical shake-up in China

David Cyranoski, Guangzhou

The bustling streets of Guangzhou — the centre of last winter's outbreak of Severe Acute Respiratory Syndrome (SARS) — no longer show much evidence of the panic that gripped the region. But the epidemic led to an array of changes in research and public health that are just beginning to make themselves felt.

This summer, plans were laid to establish a new health-research institute in the city itself. Reforms of China's biomedical research infrastructure — including the establishment of a Chinese version of the main US biomedical research agency, the National Institutes of Health (NIH) — have acquired fresh momentum. And work is under way to build or refurbish a network of laboratories with the aim of equipping them to handle dangerous pathogens.

Plans for the Guangzhou institute were confirmed on 5 July, when the Chinese Academy of Sciences (CAS), the city's government and the province of Guangdong each agreed to invest 100 million yuan (US\$12 million) in the project. The institute will be established on a new science park in western Guangzhou. A director, charged with assembling a dozen teams to do research there, will be named by the end of October.

The institute's backers say that they aim to attract a strong contingent of researchers from outside China and to establish the facility as a base for international cooperation in health research. Its areas of specialization



This incubator facility in Guangzhou's science park awaits the arrival of the new research institute.

have yet to be determined, but may include cardiovascular disease and cancer, as well as emerging diseases.

More far-reaching are proposals for a national, grant-giving biomedical research agency modelled on the NIH. The proposal was made public in a statement earlier this year by 22 senior researchers, including Zhu Chen, vice president of the CAS.

Most biomedical research in China is supported directly through branches of the academy, but advocates of the proposal want an agency that would distribute competitive grants, as the NIH does. "We need to have stable support that can build a health-research community," says Chen.

This month, a group of some 20 senior Chinese scientists worldwide, led by Ray Wu, a plant biologist at Cornell University in Ithaca, New York, and Yi Rao, an anatomist and neurobiologist at Washington University School of Medicine in St Louis, Missouri, appealed to China's prime minister Wen Jiabao to support the idea. Wu says that the group has now been asked to provide a more detailed version of its letter.

The letter criticizes what it calls China's underinvestment in biomedical research, compared with other countries. It cites the need for an organization that can carry out "a fair and transparent peer-review system". China's research structure — dominated by the CAS — has been criticized for lacking such a system.

The organization would initially distribute extramural grants, although in the future it could also support its own laboratories, as the NIH does. The NIH's annual budget of \$27 billion represents about 0.27% of the US economy, Rao says, and the proposal calls for a Chinese equivalent to spend 0.14% of China's economic product over the next ten years.

Despite Chen's support, the plan is likely to meet stiff opposition from other CAS officials, who want to retain their central position in Chinese biomedical research.

Meanwhile, many laboratories in China are pressing forward with construction of biosafety laboratories that meet the international P3 standard, which is required for most work on infectious diseases, and even the P4 standard for work on the most dangerous pathogens.

But the drive lacks coordination, researchers say. "No one can keep track of them," says Yun Zhang, a specialist at the CAS's Kunming Institute of Zoology in the province of Yunnan, which is spending 20 million yuan on labs in each class to house primate models for the study of SARS.

Secret garden opens up to public

David Cyranoski, Guangzhou

China's largest botanical garden is to get a major upgrade, as officials try to involve it in educating the public about science. But some botanists at the garden are appalled at the idea of thousands of people tramping through their precious collections.

The South China Botanical Garden in Guangzhou has until now been run primarily for research purposes. Over the next three years it will receive 300 million yuan (US\$36 million) from the Chinese Academy of Sciences, the government of Guangdong province and the city council. Plans still stress the garden's scientific role, and the project calls for 30 internationally recognized articles to be published each year. But the refurbished garden is hoped to draw two million visitors every year with its new slant on education and entertainment.

During the SARS epidemic, the number of visitors to the garden's open areas surged fivefold. "They were looking for something clean and pure," says deputy director Zhou



Pure research: botanists fear visitors seeking the purity of nature will damage the collections.

Guoyi. To keep their interest, education facilities will be expanded and the number of species increased from 5,000 to 9,000. Many additions will be wild and endangered plants, opening up new areas of research.

Nonetheless, some botanists fear that their work will be compromised and the change could harm existing collections. The most important ones — including magnolia, ginger and bamboo gardens — have until now been closed to the public. Researchers are worried about vandalism and theft.

Open-access row leads paper to shed authors

Declan Butler

A spat between the *New England Journal of Medicine* (*NEJM*) and one of the leaders of a movement for open access to the scientific literature has resulted in the journal rejecting a paper on kidney transplants at the last minute — and immediately reaccepting it without the names of four of the original authors.

Caught up in the disagreement is Minnie Sarwal, a young researcher at Stanford University School of Medicine in California, the lead author of the paper, “Molecular heterogeneity in acute renal allograft rejection identified by DNA microarray profiling” (M. Sarwal, *et al. N. Engl. J. Med.* **349**, 125–138; 2003), which was finally published on 10 July.

One of her Stanford co-authors, Patrick Brown, says he wanted the paper to be sent to an open-access journal, but reluctantly agreed to the *NEJM* as this was important for Sarwal’s career. Brown is a co-founder of the Public Library of Science (PLOS), which launches its first open-access journal next month. Papers published in PLOS journals are freely available from the time at which they are published, whereas most journals make papers available only to subscribers, for a period of time at least — six months in the case of the *NEJM*.

Brown says that he insisted that the *NEJM* publish the paper under the terms of the PLOS open-access licence, which stipulates that the authors retain copyright but agree to allow the unrestricted use, distribution and reproduction of the article in any form, provided that the original work is properly cited.

The terms on which the paper was originally accepted are now hotly disputed, however. When Brown received the galley proofs in June, a sentence — “This article is published under the terms of the PLOS open access license” — had been deleted from a previously agreed edit of the paper. Jeffrey Drazen, editor-in-chief of the *NEJM*, says that the sentence was only spotted at the last minute, and was unacceptable.

Drazen says that the offending sentence was not present in the original submitted paper, and was added later. Brown admits this, but maintains that the journal understood his terms from the outset. He had annotated the copyright form, for example, to read that submission was “entirely subject to the condition that the work in question will be made freely available for distribution and use by anyone within six months of publication”.

But Drazen is equally adamant that he would never have agreed to the PLOS licence terms, and that the *NEJM*’s editors had thought that the authors’ demands were covered by the journal’s standard policy.



Free for all: Pat Brown wants all published scientific research to be freely available.

Sarwal, who carried out the initial communications with *NEJM*, admits this possibility. “This misunderstanding may have occurred. Brown kind of let me know what sentences to write in the [covering] letter and the paper and I did so. I am sad it turned out the way it did.”

Drazen’s decision to delete the PLOS rider presented Brown with a dilemma. He says that he wanted to withdraw the paper in protest, but felt that the results were important, and should be published without further delay. Sarwal also did not want to retract the paper, as she had distributed the galleys in confidence in support of a grant proposal.

Instead, Brown called Drazen and demanded that his name and that of three other authors be withdrawn from the paper, and that this be explained in the published manuscript. Drazen refused. In a 5 June e-mail to Brown, he wrote: “We are withdrawing acceptance because all eleven original authors signed a letter, dated October

22, 2002, certifying that they were the sole authors of the work. Thus, we cannot subsequently represent to our readers that the remaining seven authors are the only authors of the entire paper.”

The paper was rejected, and a new one accepted, after Brown asked Drazen to reconsider, suggesting that the deleted authors be acknowledged as having contributed to the experiments, and sharing responsibility for the results.

Drazen says that he agreed to reaccept the paper only after Sarwal confirmed that the new authorship represented the sole authors of the work, with Brown’s team being important contributors. The paper was immediately published. Brown subsequently contacted *Nature* to persuade this journal to cover the story.

The spat has resulted in name-calling on both sides. Brown alleges that the events constitute a “clear and documented case of editorial misconduct in the handling of an article”, and that the change in authorship is “manuscript laundering”.

In a statement, the *NEJM* asserts: “It is unfortunate that Dr Brown chose to use important medical research affecting renal transplant patients to generate publicity for his planned publishing ventures. A researcher of his experience knows well that the Journal cannot selectively ignore copyright laws so that individual authors can draw attention to a personal cause. He placed his desire to promote his personal interest above his responsibility to his research colleagues.”

For her part, Sarwal says: “I am just a young scientist trying to do good science and feel terrible that any of this occurred.” ■





Dark secret: galaxy clusters such as ACO 3627 could help to unlock the mysteries of dark energy.

Cosmologists cluster to plot course towards dark energy

Geoff Brumfiel, Chicago

For five years, cosmologists and theoretical physicists have been wrestling with the bombshell discovery that a mysterious entity called 'dark energy' may be responsible for two-thirds of the energy and matter in the Universe.

Researchers are today no closer to defining dark energy, although they rely on it to explain the 1998 discovery that the Universe's expansion is accelerating, rather than slowing down under the influence of gravity, as had been widely assumed.

But they are at least closer to forging a plan to track it down. Last week, cosmologists met at the University of Chicago to hammer out a strategy to perform surveys of distant space that should glean more information about dark energy.

The plan involves constructing as many as 15 telescopes, including one in Antarctica, to work with existing ones in a bid to track thousands of clusters of galaxies. By looking at the distances between them, cosmologists hope to learn more about dark energy, which seems to make itself felt only between massive objects on a huge, intergalactic scale.

The survey will start by observing thousands of galaxy clusters up to 8 billion light years away. These clusters are observed now as they really were billions of years ago. If researchers compare the distances between these ancient clusters with those between clusters near our own galaxy — which can be observed pretty well as they are now — they may be able to establish how the clusters have been pushed apart, over time, by dark energy.

Finding and studying so many galactic clusters will be no small task, says John Carlstrom, a cosmologist at the University

of Chicago who helped to organize the conference. Researchers will need to pick out clusters from the jumble of stars and galaxies in the night sky, and make precise measurements of the microwave radiation emitted by each cluster. Microwave radiation from space can easily be distorted by atmospheric water vapour, so much of the observation will be done by telescopes in very dry locations such as Antarctica.

Previous surveys of single galaxy clusters have taken several months using existing instruments, says Carlstrom. "My group has measured about 60 clusters in a decade," he says. But with more powerful telescopes and faster data processing, the time needed should fall sharply. A 3.5-metre-diameter microwave telescope due to start observations later this year in Owen's Valley, California, will do in months what used to take years, he says.

In the long term, Carlstrom's group is also planning an \$18-million, 8-metre-diameter microwave telescope at the South Pole, which will be able to do the same survey in a few hours when it starts operations in about four years' time. Other groups are planning similar projects in the Atacama Desert in Chile and on the mountaintops of Hawaii.

But questions remain about the effectiveness of the approach. For one thing, clusters of galaxies are unwieldy and awkward to define. "A cluster isn't like a star — it doesn't have a well-defined edge," says August Evrard, a theoretical cosmologist at the University of Michigan in Ann Arbor.

Even so, Saul Perlmutter, an astronomer at the Lawrence Berkeley National Laboratory in California who led one of the groups that discovered dark energy, says the survey will represent an important step forward.

DNA lab welcomes Dalai Lama to Tibetan science community

Erika Check

Nyime Norbu has not seen his homeland, Tibet, since he and his family fled from there 30 years ago. So he was greatly moved last week when he came face-to-face with the exiled spiritual leader of Tibet, the Dalai Lama, thousands of miles from their mutual home. What did they discuss? DNA sequencing, of course.

Norbu runs DNA-sequencing machines at the Whitehead Institute/MIT Center for Genome Research, and is one of about 20 Tibetans who work there. The centre has one of the largest communities of Tibetan workers in the Boston area, which is part of the reason for the Dalai Lama's visit on 13 September during a trip to Boston.

The centre's director, Eric Lander, hired the first of the Tibetans in 1991, when the centre was preparing to sequence the mouse genome. "We needed somebody who had some scientific background, who was also very careful," he says. He ended up hiring a Tibetan refugee who was both an artist and a former inspector at a milk factory.

He in turn recruited more Tibetans throughout the 1990s, making them the most strongly represented ethnic group in the Human Genome Project: although Tibetans make up only 0.1% of the world's population, it has been estimated that they represent about 10% of the project's workforce.

"Many of our Tibetan colleagues are refugees who have learned enough science and molecular biology to become extraordinarily valuable members of our team," says Lander.

Most of the centre's Tibetans escaped the country in the face of repression by China. "To have the Dalai Lama in front of me was a blessing — we were thrilled," says Norbu. "I have no words to explain how I felt."



The Dalai Lama helps Eric Lander (right) and Yama Chopel (left) prepare mouse DNA.

United Nations set to help China plan its environmental future

London The environmental challenges associated with China's rapid economic growth are to be tackled by the United Nations Environment Programme (UNEP).

The organization's executive director, Klaus Toepfer, opened a UNEP office in Beijing on 19 September. Staff there will work with the Chinese government to tackle issues such as land degradation and water management, and will manage projects under the Global Environment Facility, an international fund that is handing out US\$3 billion for environmental projects in developing countries.



A moving issue: devastation of China's forests has forced pandas to be relocated.

China has transformed its economy over the past decade. But it has been criticized for poor environmental planning in projects such as the Three Gorges Dam, a massive hydroelectric scheme. Loss of biodiversity is also an issue: environmental organization WWF China says China is home to 156 of the world's 640 critically endangered species.

Chemists weigh up change to fundamental constant

London Chemists could soon be forced to change the textbook value of one of their fundamental constants. Avogadro's number, defined as the number of atoms in 12 grams of carbon-12, helps chemists to weigh ingredients for chemical reactions. But its tremendous size — about 6.022×10^{23} — makes it difficult to measure precisely.

The same number of atoms makes up a set weight for each individual element. Peter Becker of Germany's National Institute of Natural and Engineering Sciences in Braunschweig and colleagues measured the exact size and weight of a perfect silicon crystal to re-calculate Avogadro's number. Their result is as accurate as the previous best tally, with an error rate of one part in ten million. But the new value disagrees with the old one by one part in a million.

Chemists plan to address the issue at next year's annual meeting of the International Committee of Weights and Measures.

Japan's fraudsters face years without grants

Tokyo The Japanese government has adopted a new punishment for researchers who abuse their funding. Those caught misusing their grants-in-aid — the country's largest grant programme and lifeline of basic research — will be banned for up to five years from receiving any further money, on top of any possible fines or prison sentences.

Other countries, such as the United States, impose similar punishments for grant fraud. The move is overdue in Japan, where abuse of funds is rampant. Last month, Kamon Nitagai, then vice-president at the University of Tokyo, resigned after admitting he had submitted expense reports for false business trips, amounting to some ¥5 million (US\$44,500) from his research grant of ¥17.7 million. Nitagai says the money was used for his and his students' research.

Some researchers blame the inflexibility of the country's funding system. The strict requirement for all funding to be spent by the end of the fiscal year, for example, can encourage scientists to secretly divert money until they are ready to use it.

Ocean drilling project gets international relaunch

San Diego An international scientific programme that drills cores in the ocean floor gets a new name at the end of this month, after nearly 20 years studying the Earth and its climate.

The Ocean Drilling Program (ODP), which will be formally wrapped up at the end of September, drilled at 650 sites, taking in every ocean apart from the ice-covered Arctic. Its last trip ended at the beginning of September, when the *Resolution* returned from drilling a 2-kilometre core near Bermuda. This core will provide data from about 145 million years ago when North America and Europe were joined and the Atlantic Ocean did not yet exist.

On 1 October the ODP will be replaced by another international project, the Integrated Ocean Drilling Program (IODP). More than half of the ODP was funded by the United States, whereas the IODP's larger budget will be mainly split between the United States and Japan, which has



Sea change: the Japanese research vessel *Chikyu* will start drilling ocean cores in 2007.

contributed a new research ship, the *Chikyu*. The IODP's first project is scheduled for next June.

NASA boldly goes into the world of academia

San Diego The University of California at Santa Cruz has won a \$330-million contract to head a joint research effort with the NASA Ames Research Center.

Under the deal, signed on 15 September, university scientists will conduct mission-

specific projects for NASA in such areas as information technology, biotechnology and astrobiology. The centre will be based at NASA Ames' site at Moffett Field near San Francisco.

Europe's global navigation system heads east

London European plans to build an alternative satellite navigation system to rival the US-run Global Positioning System (GPS) received a boost last week when China agreed to back the project.

European Commission officials say that China is expected initially to invest at least E200 million (US\$230 million). Chinese officials will be trained at a new centre for satellite navigation opened on 19 September at Beijing University.

The Galileo system, which will cost more than E3 billion and is due to be operational by 2008, will be made up of 30 satellites. Its backers say the new system will also provide navigational information faster than GPS.

Correction

Abbott, A. *Nature* **425**, 4; 2003.

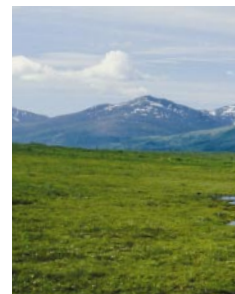
In this News article, it was wrongly stated that the cancer-vaccine trial reported in *Nature Medicine* in 2000, which has now been retracted, was supported by the healthcare company Fresenius.

Editorial policy changes *Nature* is extending its policy on competing financial interests to include the authors of Review and Progress articles. Further extensions are being considered.

Also, *Nature* has extended its policy on peer review to include considerations of risks such as assisting bioterrorism. See www.nature.com/nature/submit/policies for further details.

Too hot to handle

Alaska is warming up more than anywhere else on Earth. Climate researchers are now turning to regional models to find out why — and how to deal with it. John Whitfield went north to investigate.



D. BELTRA/REUTERS

The vast forests that cover southern Alaska should be evergreen. But not these days. Hop in a tiny plane — which Alaskans seem to do as often as a New Yorker hails a cab — and you'll see patches of brown stretching for miles into the wilderness. This is the work of the spruce bark beetle, which over the past 15 years has killed more trees in Alaska than any other insect in North America's recorded history. In the Kenai Peninsula on Alaska's southern coast, some 40 million spruce have perished across an area twice the size of Yellowstone National Park. The beetle's population rocketed thanks to changes in the weather, argues Ed Berg, an ecologist with the Kenai National Wildlife Refuge. "We had a really long run of warm summers," he says.

The beetle boom is one of the more dramatic changes that locals and scientists alike attribute to global warming. But it's just one in a long list. Farther down the coast in Prince William Sound, boats pick their way through increased numbers of icebergs calving off the Columbia Glacier. The glacier has retreated 12 kilometres over the past 20 years, and some say it could collapse completely in another ten. Alaska's notorious mosquitoes, so big they're jokingly referred to as the state bird, have spread north to irritate the few residents who used to escape their attentions. Even the plants are changing. The spongy tundra, usually covered in grass and moss, is slowly being invaded by woody shrubs¹.

This summer saw the biggest melt yet in Alaska's sea ice, and winter in the interior was unprecedentedly mild — for the Arctic. "I've lived here since 1968, and last winter was the first one that didn't drop below -40°C ," says Gunter Weller, director of the Center for Global Change and Arctic System Research at the University of Alaska, Fairbanks.

Small change, big difference

Temperatures have changed more in Alaska over the past 30 years than they have anywhere else on Earth: winters have warmed by a startling $2-3^{\circ}\text{C}$, compared with a global average of 1°C . That's guaranteed to have dramatic effects in an Arctic landscape, where even small temperature changes can make the difference between freezing and melting. In Fairbanks, a city built on permafrost, the annual mean temperature is just -2°C . If it pops above zero, residents can say goodbye to the frozen ground beneath



their feet, along with the free iceboxes in their basements. The impacts on wildlife, and the people who depend on it for their livelihoods, will be huge.

Determining the effects of climate change on a local level has become one of the major priorities of the Intergovernmental Panel for Climate Change². And to do that, the panel needs regional models. Unlike the first generation of climate models, which covered the whole globe and divided it into chunks 50–200 kilometres a side, small-scale, regional models now look at sections just a few kilometres across. This helps to refine the picture of climate change and lets models take account of elements of topography such as mountain ranges, peninsulas and even cracks in sea ice.

Such models are particularly crucial for places such as Alaska, as global-scale simulations are notoriously inaccurate at the poles³. Arctic clouds, for example, are not like their counterparts at lower latitudes: their droplets may be mostly ice rather than liquid water, for example. The gaps in our knowledge about arctic clouds puts some rainfall predictions off by as much as 100%. And no one can agree on how to deal with the reflectance of the Sun's energy by ice and snow, known as albedo.

"Small differences in the value of ice albedo can produce large differences in model outputs," says climate researcher Filippo Giorgi of the Abdus Salam International Center for Theoretical Physics in Trieste, Italy.

Polar confusion

Inaccuracies in modelling other parts of the global climate can also accumulate up at the poles: an error in describing the winds that carry heat from the tropics to the poles, for example, could increase with every degree of latitude. "The Arctic could be a dumping ground for errors elsewhere," says atmospheric scientist Judy Curry of the Georgia Institute of Technology in Atlanta.

It is also hard to tell how much of Alaska's climate change is due to global warming and how much to natural climate cycles. The Pacific Decadal Oscillation — an El Niño-like fluctuation of temperatures between the north and tropical Pacific that takes place over 20–30 years — flipped Alaska into a warming phase in the 1970s. The North Atlantic Oscillation has also contributed to warmer winters in Alaska since the late 1960s⁴. But at the same time it has been associated with a $2-3^{\circ}\text{C}$ cooling just across the continent in Greenland. There seems to be a 60-year see-saw in

A tiny rise in Alaska's average temperature has let beetles wreak havoc in the forests (bottom left) and turned permafrost into marsh (left). Early spring thaws spell bad news for wildlife and trap Inuit hunters out on the snow (below).

Those model results can't come soon enough for the people of Alaska. In the far north, the area covered by sea ice is shrinking at a rate of about 3% per decade⁶ — bad news for the seals and polar bears that depend on the ice environment, and for the subsistence hunters who depend on the animals. Winter is the prime hunting season, as snowmobiles can cover distances much more quickly than any vehicle on the tundra of summer. But the date when the snow melts has become less predictable, forcing communities to invest in expensive contingency plans such as helicopter rescue for stranded hunting parties.

Slip sliding away

The inhabitants of Barrow, the northernmost town in the United States, stand to lose more than their meals. The retreat of sea ice has exposed the land to the sea for more of the year, which has meant more erosion of the coast. Other symptoms of climate change exacerbate the erosion: more storms, melting permafrost — which makes the ground softer — and higher sea levels. Several other Alaskan communities are faced with having the ground washed out from under them, too. There are even plans to dismantle and move two villages on the northwest coast, Shishmaref and Kivalina, at a cost of more than \$100 million — over \$100,000 per resident.

Lynch and her colleagues have plotted the course of erosion in Barrow and the efficacy of attempts to stop it, such as building concrete walls or reinforcing beaches with sand. They have concluded that costly sea defences have done little to stem erosion. But they have also found that some areas of Barrow's coastline are actually gaining new land from the sea. When new buildings are required, says Lynch, the best response to the threat of erosion would be to focus construction in these places. Better prediction of future erosion will be the big challenge for her group over the next few years, she says.

As the ARC-MIP models are improved, they will help to refine global models, in preparation for a time when there is enough computer power to give global models the same resolution as local ones. Alaska should help to inform the global picture in other ways, too. Whether Alaska's current bout of warming slows down or even reverses, the dramatic effects of a few degrees of warming seen there should put us all on alert. "What you see here is an indicator of what can be expected in the rest of the world," says Weller. For much of the planet, that means dead trees, mosquitoes and a lot of slush. ■

John Whitfield works in Nature's news syndication team.

1. Sturm, M., Racine, C. & Tape, K. *Nature* **411**, 546–547 (2001).
2. Giles, J. *Nature* **417**, 106 (2002).
3. Walsh, J. E., Kattsov, V. M., Chapman, W. L., Govorkova V. & Pavlova, T. J. *Clim.* **15**, 1429–1446 (2002).
4. Thompson, D. W. J. & Wallace, J. M. *Science* **293**, 85–89 (2001).
5. Walsh, J. E., Lynch, A. H., Chapman, W. L. & Musgrave, D. *Meteorol. Atmos. Phys.* **51**, 179–194 (1993).
6. Vinnikov, K. et al. *Science* **286**, 1934–1937 (1999).

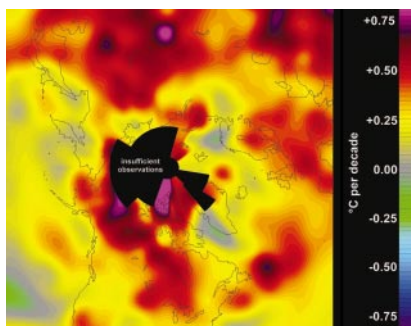
temperatures between the east and west of this part of the Arctic: when one side heats up, the other cools down. In a few years, this could flip again and reverse Alaska's recent warming trend. No one yet knows.

So it is perhaps not surprising that Alaska was one of the first places to attract the serious attention of regional modellers. Back in 1993, Amanda Lynch of the University of Colorado, Boulder, created the world's first detailed regional model to take into account such interactions as those between ice and air and water, and she did it for Alaska⁵. Since then, Lynch has refined her model and been

joined by a host of teams aiming to pin down Alaska's climate.

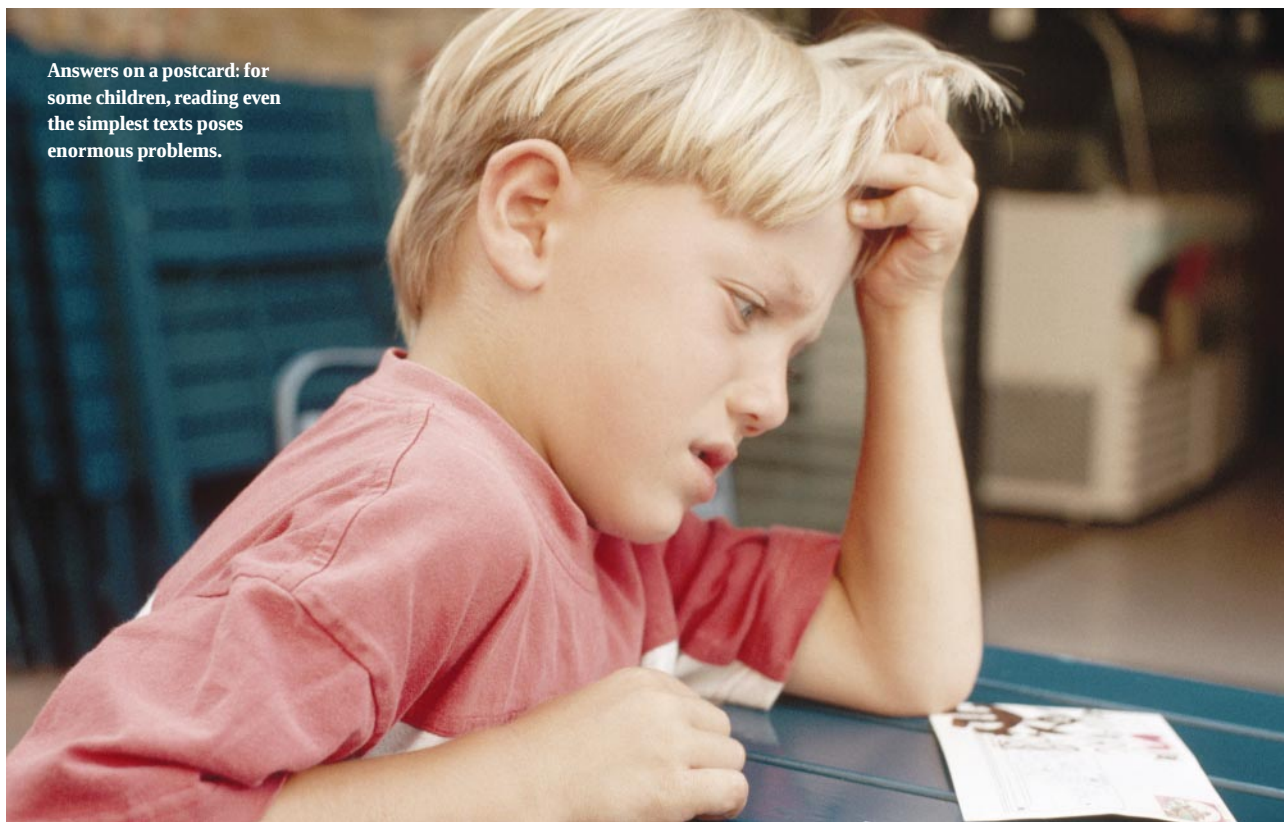
In 2000, groups from the United States, Canada, Germany and Sweden, all with their own regional models, teamed up to run the Arctic Regional Climate Model Intercomparison Project (ARC-MIP). It isn't a beauty contest, says Curry, who serves as the project's coordinator. "The goal isn't to identify the best model, it's to improve all models," she says. The main difference between participants is how they deal with the interface where air meets sea or land: some have very complex treatments of how vegetation or sea ice influence the flow of gas and water into and out of the atmosphere. Others are less complex, treating sea ice as a simple barrier.

The first trial set for ARC-MIP's models has been to reproduce the results from one of the best climate data sets for the poorly surveyed Arctic — measurements collected by an icebreaker that was frozen into Arctic sea ice from October 1997 to October 1998. Those first ARC-MIP results should be submitted to journals around the end of this year, says Curry. The project is far from over, but so far it looks as if the simpler models are performing more accurately: complexity may just create more opportunities for things to go wrong.



Heating up: Alaska is a hotspot in this plot of air temperature change from 1973 to 2002.

Answers on a postcard: for some children, reading even the simplest texts poses enormous problems.



Lost for words

Thanks in part to brain-imaging technology, researchers are now homing in on the root cause of dyslexia. But research into strategies for treating the condition is still in its infancy, says Glenn Murphy.

Most people have a vague idea of what it is, or know someone who has it, or have heard about famous sufferers such as actor Tom Cruise, who seem to get by well enough despite their problems with reading. But for most of the past century, researchers have been unable to agree on what causes dyslexia.

When children or adults fail to learn to read fluently — despite normal intelligence, instruction and opportunities to do so — they are diagnosed with developmental dyslexia. Exactly why they should fail has led to speculation about a host of possible causes. Over the past few years, however, brain-imaging studies have supplied fresh evidence that the fundamental problem lies in the brain's ability to process 'phonemes'. These are the speech sounds that enable us to tell one word from another — 'pet' and 'bet', for instance, are distinguished by the sounds of their initial consonants.

For all of us, decoding the arbitrary sym-

bols of written language is a complex skill that requires instruction and considerable effort to attain. And for the world's dyslexics — around one in ten people — it presents

"My parents found out I was dyslexic when I was five. My aunt was training to be an educational psychologist, and she pointed out that something wasn't quite right."

Jess, 25, from London

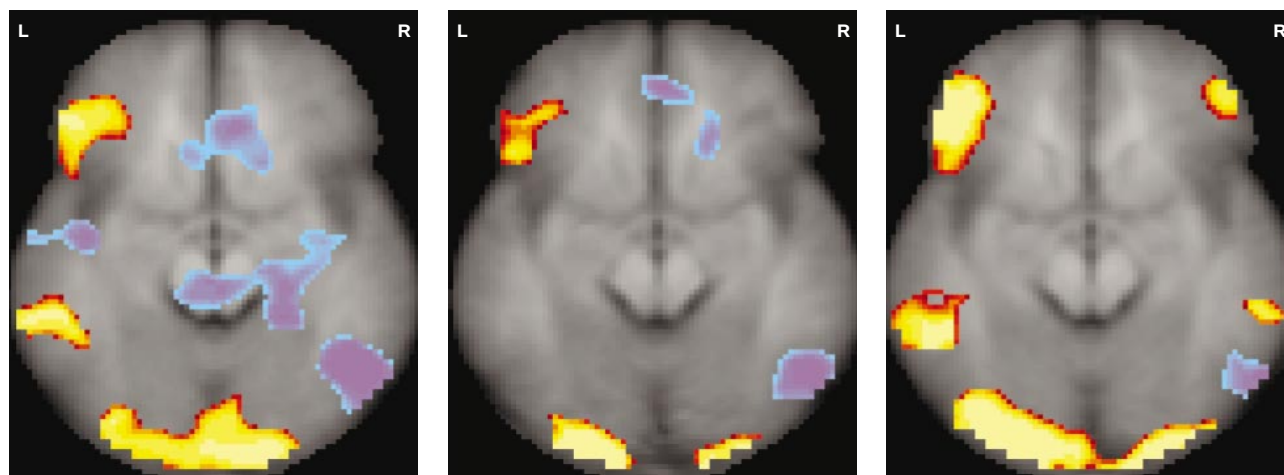
almost insurmountable problems. Their plight first attracted attention in the late nineteenth century. Then studied mostly by ophthalmologists, speculation about the cause of this 'word blindness' centred initially on problems with vision — a view that still has its adherents today. In the 1980s, for

instance, physiologist John Stein of the University of Oxford, UK, developed a theory that blames dyslexia on problems with focusing on text and with the scanning eye movements that we perform while reading¹.

Other researchers have argued that the underlying problem lies with auditory processing. In 1980, Paula Tallal at Rutgers University's campus in Newark, New Jersey, proposed that dyslexics suffer from a defect that hampers their perception of short or rapidly varying sounds². The defect is

not disruptive enough to prevent dyslexics from learning to understand speech, but Tallal argues that it prevents them from associating short bursts of phonemes with their respective letters as they are learning to read. Some dyslexics also have problems with movement and balance, which has led psychologist Roderick Nicholson of the University of Sheffield, UK, to argue that dyslexia is caused by defects in the cerebellum, at the base of the brain³.

In recent years, these ideas about visual, auditory and cerebellar defects have become part of a larger, 'magnocellular' theory of dyslexia⁴. This stems from post-mortem evidence — provided in 1991 by Margaret Livingstone and Albert Galaburda of Harvard Medical School in Boston⁵ — showing that dyslexics have abnormalities in the magnocellular visual pathway, which is involved in processing fast-changing visual information. Other magnocellular systems handle similarly fleeting auditory and tactile information, and they each feed into the cerebellum during learning — so an extensive magnocellular dysfunction could explain



During reading, persistently poor readers (right) use similar brain areas to normal subjects (left), most notably in the back of the brain (bottom of images), although with markedly less success. Dyslexics who have to some extent overcome their disability (middle) show a different pattern of activity.

many of the defects observed in dyslexics.

Many dyslexia researchers, however, belong to a rival camp that argues that the fundamental problem lies not in visual or auditory processing, but is instead caused by a more specific defect in the brain's ability to decode phonemes. In tests, for example, dyslexics might swap the first letters of a pair of arbitrarily chosen words — turning 'basket-lemon' into 'lasket-bemon', for instance.

Over the past two decades, supporters of this phonological theory have amassed evidence documenting dyslexics' problems with phoneme processing, and confirming that these defects are the most commonly observed symptom of dyslexia⁶. In an attempt to close the case, Uta Frith of the Institute of Cognitive Neuroscience at University College London (UCL) and her postdoc Franck Ramus have recently completed the first study to test the competing theories of dyslexia in the same experimental subjects.

Frith, Ramus and their colleagues subjected 16 dyslexic volunteers and an equal number of non-dyslexic controls — all students at UCL — to a barrage of tests of hearing, vision, balance, motor coordination, intelligence and phoneme awareness. "Each person got at least ten hours of testing," explains Frith, "using tests that the proponents of the various theories themselves invented and suggested."

The results, published in April, revealed that all of the dyslexics had phonological deficits, and five showed none of the symptoms implicated in the rival theories⁷. "While only some dyslexics have abnormal vision and hearing, all have problems with tasks that specifically require them to manipulate phonemes," says Frith. This, she argues, indicates that dyslexia is essentially a disorder of phoneme processing; visual,

hearing and cerebellar problems may often be associated with the condition, but they are not its direct cause.

Supporters of the magnocellular theory maintain that poor phonological processing is just one of many symptoms of dyslexia. But most experts now see it as the defining symptom, and the phonological theory has come to dominate. "There is more and more supporting evidence piling up for the phonological model, and conflicting evidence piling up against the others," says Maggie Snowling, a dyslexia researcher at the University of York, UK.

Some of the most persuasive evidence comes from studies using functional magnetic resonance imaging (fMRI), which allows researchers to study activity in the brain while it performs specific tasks, by monitoring changes in blood flow. "Functional imaging is a wonderful tool," says neuroscientist Sally Shaywitz of Yale University in New Haven, Connecticut. "It allows you to use fewer subjects and target different neural systems."

Results from fMRI suggest that there are at least two pathways for reading in the brain: inexperienced readers use one pathway, whereas a second, faster pathway takes over in more skilled readers. Both involve three key areas in the left side of the brain: a region at the front of the brain known as Broca's area; and at the rear, the parieto-temporal and occipito-temporal regions (see diagram, above). Broca's area has long been known, from studies of patients with brain lesions, to be required for normal speech and writing. Novice readers

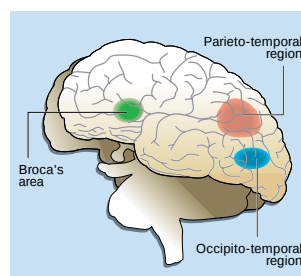
seem to use the parieto-temporal region to dismantle words for step-by-step phonological analysis; more experienced readers apparently rely on the occipito-temporal region to recognize whole words instantly⁸.

With her husband Bennett and other Yale colleagues, Shaywitz has used fMRI to compare the brains of dyslexics to those of normal, healthy readers as they perform reading tasks such as trying to identify nonsensical words in rhyming pairs and real words in non-rhyming pairs. When the words used were nonsensical ('jeat' and 'lete', for example), they could not instantly be recognized, and so all the volunteers were forced to sound out the words in their heads, phoneme by phoneme. In 1998, Shaywitz and her colleagues reported that dyslexics and non-dyslexics differ in their patterns of brain activity⁹. "We found a clear neurological signature for dyslexia," she says.

As expected, the brains of normal readers lit up on the left side, activating areas mostly at the back, in particular the parieto-temporal region. But dyslexics showed a different pattern: the back of the brain remained dim in the scans, whereas the front, around Broca's area, was highly active. The dyslexic volunteers were apparently trying to compensate for

their inactive parieto-temporal region by overactivating other parts of the brain's reading circuitry.

The dyslexic volunteers in this study were adults, which left open the possibility that their distinctive patterns of brain activity were the consequence of years of struggling with reading, rather than being the cause of their problems. But Shaywitz has since repeated the work with 144 children — half of them dyslexic, the rest normal readers.



Map reading: the brain regions that are involved in deciphering text.

Last year, her team reported that the telltale fMRI signature had appeared once again⁸, bolstering the idea that the underactivation of the parieto-temporal region seen in the brain scans is a key component of the underlying neurological deficit that causes dyslexia.

Shaywitz's research has also yielded insights into the processes by which some dyslexics are able to overcome, at least partially, their difficulties with reading. Although the fMRI scans of Shaywitz's older and younger dyslexic volunteers were broadly similar, there were some subtle differences: the overactivation of the front of the brain was more pronounced in older readers, who also activated regions on the right side of their brains. Could this mean that the older dyslexics were able to offset their faulty phonological processing systems by using other pathways?

To investigate the issue more thoroughly, Shaywitz looked at a group of young dyslexics who had managed to become accurate, if not fluent, readers, and compared their brain activity with that of another group whose reading had remained persistently poor. The two groups had been monitored since they were five years old, and assessed annually as part of a wider study of the acquisition of reading skills.

Shaywitz examined these young adults using fMRI, and again discovered distinct patterns of brain activity in each group. Those who could partially compensate for their reading problems had similar fMRI patterns to the dyslexic signatures identified in Shaywitz's earlier studies, with reduced activity at the rear of the brain. Persistently poor readers also gave similar results for tests involving nonsensical words. But when tested with real words, their brains became active in the occipito-temporal region. This was different from the pattern of activity in this area in normal readers, and seemed to be connected to activity in areas that are involved in general visual memory, rather than those that deal specifically with reading and language¹⁰.

Shaywitz argues that many dyslexics learn to compensate for their poor phonological processing ability by increasing the activity of Broca's area. The persistently poor readers, on the other hand, try to rely more on general rote memory.

For researchers who are devising methods to help dyslexic children learn to read,



Word processor: the Fast ForWord program helps dyslexics associate sounds with letters.

"As you get older, you try to make an effort to improve, and you do. I'm nowhere near as bad as I used to be, because I have to make a conscious effort to get things right—for fear of looking daft at work."

Jess

this is an important result. First, it suggests that there may be at least two subgroups of dyslexics, one of which will respond more effectively to treatment. Second, it provides hope that the new consensus on the importance of phoneme processing, combined with the ability of functional brain imaging to monitor the activity of dyslexics' brains as they learn to overcome their difficulties, will help to put research into treatment strategies on a sounder scientific footing.

One study has already used fMRI to help monitor the effectiveness of a popular computer program used to treat dyslexia. Fast ForWord, developed by Tallal in collaboration with John Gabrieli of Stanford University in California, was devised in the light of Tallal's auditory theory of dyslexia, and therefore focuses on helping dyslexic children to process speech by exaggerating words and slowing them down.

Tallal and Gabrieli trained 20 dyslexic children, aged between 8 and 12 years, for eight weeks, giving them 100 minutes of tuition with Fast ForWord each day. In February, the researchers reported that the children's reading skills were significantly improved, and that these changes correlated with visible changes in brain function, viewed by fMRI as the children tackled a series of rhyming exercises¹¹. There were "huge differences", says Gabrieli, including activation of the brain areas deployed by normal readers, and of the areas that are implicated in compensation for dyslexia by Shaywitz.

It's an encouraging result, but one that should be greeted cautiously, says Guinevere Eden of the Center for the Study of Learning

at Georgetown University in Washington DC. She points out that Tallal and Gabrieli's study omitted a necessary control group, meaning that the placebo effect cannot be ruled out. "You need to involve another group of dyslexics that don't get the treatment," Eden says.

This sort of problem is common in research into dyslexia treatment, she says, as even studies with appropriate control groups have failed to match dyslexic and non-dyslexic subjects for general intelligence, socio-economic status and other factors. Eden also argues that more work will need to be done on the neurological basis of ordinary reading before improved treatment methods can be developed for dyslexia. "We're operating in a vacuum," she says. Functional brain-imaging studies are improving our understanding, she adds, but a great deal more groundwork remains to be done before the results can be integrated into education programmes¹².

Another difficulty for the field is that many current treatment strategies are linked to commercial products such as the Fast ForWord program. In this business-oriented climate, it is difficult for researchers, who may have a stake in the products they are testing,

to remain objective, Eden says.

But the stage should now be set for more rigorous studies using brain imaging and other tests to evaluate the success of the various products on the market — especially those that focus specifically on improving phonological processing — and to help devise better treatment methods. "The big news," says Shaywitz, "is that science has come into education. It's amazing that it has taken so long to do so."

Glenn Murphy is a student in science communication at Imperial College, London.

1. Stein, J. F. & Fowler, S. Br. J. Ophthalmol. **66**, 332–336 (1982).
2. Tallal, P. Brain Lang. **9**, 182–198 (1980).
3. Nicholson, R. I. et al. Lancet **353**, 1662–1667 (1999).
4. Stein, J. F. Dyslexia **7**, 12–36 (2001).
5. Livingstone, M. S., Rosen, G. D., Drislane, F. W. & Galaburda, A. M. Proc. Natl Acad. Sci. USA **88**, 7943–7947 (1991).
6. Ramus, F. Curr. Opin. Neurobiol. **13**, 212–218 (2003).
7. Ramus, F. et al. Brain **126**, 841–865 (2003).
8. Shaywitz, B. A. et al. Biol. Psychiatry **52**, 101–110 (2002).
9. Shaywitz, S. E. et al. Proc. Natl Acad. Sci. USA **95**, 2636–2641 (1998).
10. Shaywitz, S. E. et al. Biol. Psychiatry **54**, 25–33 (2003).
11. Temple, E. et al. Proc. Natl Acad. Sci. USA **100**, 2860–2865 (2003).
12. Eden, G. F. & Moats, L. Nature Neurosci. **5**, 1080–1084 (2002).

Thanks to Jess for the insights into life as a dyslexic quoted through this article.

Global project needed to tackle coffee crisis

A sharp drop in coffee prices has caused widespread suffering and hindered research.

Sir — The importance of coffee as an agricultural commodity cannot be overstated: its retail value of US\$70 billion surpasses the forecast of \$56 billion for total US agricultural exports for 2003. Although coffee is the world's most heavily traded commodity apart from oil, it has been overproduced for several years: some 117 million 60-kg bags were produced in 2002–2003 but only 108 million were consumed. Overproduction has resulted in historically low coffee prices (adjusted for inflation) of about 50 US cents a pound — the measure in which it is sold in international markets — or \$1.10 a kilogram. Producing countries received about \$5.5 billion out of the \$70-billion total retail value for 2002–2003, compared with \$10–12 billion received out of the \$30-billion retail value in the early 1990s.

Low prices are having a devastating effect on at least 20 million coffee-farming families in more than 50 countries. Taking an average family size of five, more than 100 million people are dependent on coffee. In Ethiopia, for example, more than 700,000 families are involved in coffee production and more

than 15 million people depend on coffee directly or indirectly. In Central America, 540,000 coffee workers lost their jobs between 2000 and 2002 because of low coffee prices. Efforts to reduce poverty are being seriously hampered by such a widespread crisis.

One rarely acknowledged victim of the coffee crisis is research, for which funding has been reduced. At Cenicafe, the National Coffee Research Centre in Colombia, for example, the workforce was reduced from 436 in 1988 to 169 in 2001, with further cuts in 2002. Research on coffee harvesting, processing, pest and disease control, post-harvest storage and mycotoxins, among others, is essential for the continued production of high-quality coffee. It is imperative for coffee research to be maintained, yet such efforts appear to be lost in the current economic crisis.

One mechanism for maintaining continuity in coffee research globally would be the creation of an international coffee-research development programme within one of the 16 centres operating under the auspices of the Consultative Group on International Agricultural

Research (CGIAR). A similar programme has already been set up for banana and plantain within the International Plant Genetic Resources Institute. A coffee equivalent would coordinate global efforts in research, training, data-gathering and database management, and would report results, serving all coffee-producing countries. Global research priorities could be coordinated through the programme, minimizing duplicated research and making more efficient use of scarce coffee-research funds. The programme should also aim to facilitate development of research-based solutions to major production and quality problems, with smallholders in developing countries as the primary focus. Its structure should be non-bureaucratic, rapid-response, flexible and adaptable.

Fernando E. Vega*, **Eric Rosenquist†**,
Wanda Collins‡

**Insect Biocontrol Laboratory, US Department of Agriculture, Agricultural Research Service, Beltsville, Maryland 20705, USA*

†National Program Staff, USDA, ARS, Beltsville, Maryland 20705, USA

‡Plant Sciences Institute, USDA, ARS, Beltsville, Maryland 20705, USA

The public has its own view of what is a risk

Sir — The assertion in your Editorial “Don't believe the hype” (*Nature* **424**, 237; 2003), that new toxicology data on nanomaterials and more public debate on their risks will extinguish the hype surrounding nanotechnology, is unlikely to be correct. Lessons learnt from past controversies suggest that what is actually needed is understanding of what the public perceives to be the risks of nanotechnology.

Collection of scientific data and facts on potential hazards does little to calm the fears of the public, as experiences of silicone breast implants and *Bt* corn demonstrate. Even scientific evidence showing that there is little or no harm from a particular technology does not put an end to public controversy. It is public attitudes and reactions to perceived, not actual, risks that tip the balance of public acceptance or rejection of new technologies. Why is this so?

Statistics about risks — as assessed by scientific methods — do not necessarily guide the public towards rational decisions (P. Slovic, *Science* **236**, 280–285;

1987). Slovic and his colleagues have shown over the past three decades that non-technical people tend to overestimate the risks of activities that are unfamiliar to them, that may threaten future generations, and that have vivid historical associations, such as Three Mile Island or Chernobyl.

The nanotechnology community is underestimating the importance of perceived risks — and the distinction between these and real risks — and their implications for progress.

In 2002, the US National Nanotechnology Initiative awarded only \$280,000 — 0.04% of its budget of \$697 million, to study the social and ethical implications of nanotechnology. None of this money was allocated to studying risk perception.

The longer the nanotechnology community waits to address public concerns, the more entrenched risk perception will become in the public's minds. It is up to us to take notice and assess funds accordingly.

Emmanuelle Schuler
Center for Nanoscale Science and Technology, Rice University, 6100 Main Street, Houston, Texas 77005-1892, USA

Feynman put a personal spin on physics

Sir — Thomas Halsey comments, in his News and Views article “Friction in a spin” (*Nature* **424**, 1005; 2003), “No doubt string theorists and creators of Bose–Einstein condensates will be bemused to discover that they are sharing academic departments with colleagues whose idea of fundamental physics involves spinning coins.”

Or perhaps not, if history is anything to go by. The physicist Richard Feynman famously put on record how one day in the cafeteria “some guy, fooling around, throws a plate in the air”. By noticing the difference between the plate's angular velocity and that of the associated wobble, says Feynman, he was motivated to higher things: “The diagrams and the whole business that I got the Nobel Prize for came from that piddling around with the wobbling plate” (R. P. Feynman with R. Leighton *Surely You're Joking, Mr Feynman!* Norton, New York, 1985).

R. W. D. Nickalls
Department of Anaesthesia, Nottingham City Hospital, Hucknall Road, Nottingham NG5 1PB, UK

Learning to evolve

A fresh look at whether learning alters the course of natural selection.

Evolution and Learning: The Baldwin Effect Reconsidered

edited by Bruce H. Weber &

David J. Depew

MIT Press: 2003. 352 pp. \$45, £29.50

Kevin N. Laland

At the start of the twentieth century, James Mark Baldwin, professor of psychology at Johns Hopkins University and former president of the American Psychological Association, was one of the United States' leading developmental psychologists. Tragically, he was forced to resign his posts after being arrested in a brothel, and his substantial contributions to psychology were written out of the history books.

This salacious snippet would probably have been long forgotten had Baldwin not left the conceptual legacy of a mechanism purporting to allow 'mind' to guide evolution. The idea is that organisms could survive ecological challenges by virtue of their acquired knowledge and skills, frequently learned from others, and that this would then channel natural selection to favour unlearned versions of the same adaptive behaviour, perhaps because such variants more reliably facilitate survival. Conwy Lloyd Morgan and Henry F. Osborn published identical conjectures at about the same time, but Baldwin asserted his priority, and the mechanism is now known as the Baldwin effect.

Baldwin received heavyweight backing from Julian Huxley, Jean Piaget and Conrad Waddington among others, but few evolutionary biologists have allotted much thought or credence to the notion, and leading figures such as George Simpson and Ernst Mayr were hostile. But the Baldwin effect has recently re-emerged, receiving theoretical support, being endorsed by John Maynard Smith and cropping up in theories of the evolution of language and mind.

In this collection of thought-provoking essays on learning and evolution, editors Bruce Weber and David Depew set out to reconsider the Baldwin effect in the light of these developments, exploring whether it has anything to add to contemporary debates about evolution. The 14 contributors are mainly philosophers, psychologists or developmental biologists; strikingly, none are evolutionary biologists.

Although not all are 'Baldwin boosters', most concur that there is something missing from modern evolutionary biology. For instance, Terrence Deacon and Daniel Dennett feel the need for tools in addition to natural selection to describe their theories



Cheese spread: the emergence of dairy farming may have created the selection pressures that favoured genes for lactose absorption.

of language and brain evolution, and evoke the Baldwin effect as the best available mechanism. Celia Moore, Susan Oyama and Paul Griffiths would like to see a broadening of the notion of inheritance to include epigenetic and social factors. Highlights of the collection include Deacon's imaginative account of the coevolution of language and brain, a delightful essay by Moore depicting the complexities of behavioural development with vivid examples, and a characteristically astute deconstruction of the Baldwin effect by Peter Godfrey-Smith. But the contributor who puts his finger on the problem is Griffiths.

Baldwin was a prescient character who anticipated much of the causal logic behind gene-culture coevolutionary theory and developmental-systems theory. And he was right. For example, around the world, both the proportion of the adult population able to consume dairy products without becoming sick and the frequency of genes for lactose absorption covary strongly with a history of dairy farming. Recent comparative analyses reveal that dairy farming emerged before the spread of genes for lactose absorption and almost certainly created the selection pressures that favoured them, not the other way around.

Yet, as Griffiths argues, the Baldwin effect distracts attention from the more fundamental processes of niche construction — the idea

that organisms shape the environments in which they live — and ecological (extragenetic) inheritance. The only sense in which the Baldwin effect can be regarded as a new factor in evolution is that traits not fully specified by naturally selected genes modify selection pressures acting back on the organism. There is no reason why such traits should be restricted to learned behaviour, because few traits are genetically determined, nor why the modified selection should be restricted to that favouring genes for the trait. All living creatures, through their metabolism, activities and choices, partly create and partly destroy their own niches.

I suspect that evolutionary biologists feel that they can ignore Baldwin because the early developmental sequestering of germline from somatic cells means that developmental processes, including learning, do not directly affect genetic inheritance. Yet by modifying selection pressures that act back on the organism, niche construction creates an indirect causal pathway. The widespread assumption that niche construction is a product of evolution but not a process in its own right is tantamount to the assumption that development is fully determined by naturally selected genes.

A minor gripe about the book is that it is sometimes repetitive: after the first couple of chapters the reader no longer needs to be told what the Baldwin effect is. Furthermore, some relevant empirical and theoretical findings are ignored, including much of the recent work on the evolution of plasticity and reaction norms, and the best theoretical analysis of the Baldwin effect (L. W. Ance, *J. Theor. Biol.* **196**, 197–209; 1999). Yet these deficiencies detract little from the exercise. *Evolution and Learning* is a readable and challenging volume, and I would recommend it strongly to people who enjoy thinking hard about evolution. ■

Kevin N. Laland is at the School of Biology, Bute Medical Building, University of St Andrews, St Andrews, Fife KY16 9TS, UK. He is a co-author of *Niche Construction: The Neglected Process in Evolution* (Princeton University Press, 2003).

Biologists become clock-watchers

Chronobiology: Biological Timekeeping

edited by Jay C. Dunlap, Jennifer J. Loros & Patricia J. DeCoursey
 Sinauer: 2003. 382 pp. \$74.95
 Palgrave: 2003. £47.99

John Palmer

It is well known that the lives of most plants and animals are tuned to the night-day cycle. Most birds are active during the day, most bats at night. Many flowers open during the day and close their petals at night. Except for insomniacs and shift-workers, humans sleep at night. For centuries the axiomatic interpretation was that these activities were directly governed by days alternating with nights. But halfway through the twentieth century this interpretation was found to be inadequate; it was discovered then that within the bodies of virtually all organisms is a living clock that governs their periodic behaviour. The discipline of chronobiology has since been created to study this innate periodicity, the ultimate goal being to decipher the clockwork.

Chronobiology was written to support a course on this subject for advanced undergraduates and graduate students in the life sciences. At

present, no other book is well suited for such a course. Its opening chapter provides an overview of the field, but I feel that many readers will have problems in understanding it. It presents too much, too briefly (in just 12 pages), before uninitiated readers have even been given a feeling for what biological rhythms are. That could be a turn-off to some.

The information presented in the second chapter might have made a better introduction. It introduces the reader to the most commonly studied clock-controlled rhythms in the plant and animal kingdoms, emphasizing those based on the 24-hour length of a day. It then describes the clock's use in Sun-compass orientation (it corrects for the changing position of the Sun as it crosses the sky), seasonality and time sense. The examples are well-chosen and informative. The book then goes on to describe how rhythms are entrained to environmental cycles, and discusses the innate nature of the clock and the clockwork's general rate-insensitivity to different constant temperatures.

Next up is a section about the genetics of the clock (the *frq* gene in the fungus *Neurospora*, and the *per* gene in the fruitfly *Drosophila*) and the location of the clock in various organisms (the brain in certain insects, the eye in some molluscs, and the eye, pineal gland and part of the brain known as the suprachiasmatic nuclei in vertebrates). There is a discussion about how these clocks may be connected to physiological and behavioural processes, causing them to be rhythmic. Then there are two chapters covering the various aspects of rhythms in humans, followed by a concluding account suggesting the directions of chronobiological research in the near future.

On a first quick pass through the book I noticed that each chapter was a construct of many contributors — about 70 in all, mostly leading chronobiology researchers. And this immediately brought to mind the adage that the curious anatomy of a giraffe signified that it must have been designed by a committee. So I began reading with some trepidation. But my

concern was for nought. The presentation reads smoothly (although it is wordy and too academic in places) and is well-organized — about half of it was written and overseen by one of the editors. All of the topics that must be covered in a text on chronobiology are present. And the production of the book is superb: the many illustrations and photographs are beautifully reproduced and the layout is uncluttered and appealing, so students should like it.

In a nutshell, I take my hat off to the many contributors and the publisher for their efforts. The stated goal of the preface has been accomplished — the field now has its much-needed textbook. ■

John D. Palmer is in the Department of Biology, University of Massachusetts, Amherst, Massachusetts 01003, USA.

A dangerous world?

The Suffering Gene: Environmental Threats to our Health

by Roy Burdon

Zed Books: 2003. 227 pp. \$59.95, £45 (hbk); \$19.95, £12.99 (pbk)

Roger Cox

In *The Suffering Gene*, Roy Burdon explores areas of environmental concern in which science, special interests, public opinion and regulatory requirements are frequently uncomfortable bedfellows. He addresses the extent to which environmental factors, particularly those associated with industrial activities and new technologies, can affect human health. The book also touches on the difficulties of establishing an acceptable balance between the socio-economic benefits of certain human activities and the associated risks to health.

Burdon sets about his task with obvious enthusiasm. He commences with outlines of DNA structure, function and response to damage, and subsequently covers the known or suspected health effects of different types of radiation, chemical agents and free radicals. The biological mechanisms that operate to defend against these health effects (such as DNA repair and apoptosis) also receive attention. The main emphasis of the early chapters is the risk of cancer, including heritable susceptibility, but germline mutations, developmental abnormalities and ageing are also addressed.

Later chapters cover methodologies and concerns relating to recombinant organisms and human cloning. Burdon also includes discussion of the possible consequences of global climate change and, at various points in the book, touches on topics that are somewhat peripheral to the main environmental theme, such as cancer therapy and gene therapy to treat inherited diseases.



Linnaeus showed that you can set your clock by different flowers' opening times.

All of these topics are dealt with in 227 pages, so this is not a book for specialist readers — many of whom will find, as I did, some deficiencies in the scientific discussions and a few errors of fact or interpretation. However, this broad coverage, together with Burdon's ability to write in a straightforward and engaging manner, should ensure that a lay audience will find it a useful scientific introduction to the effects of the environment on health.

For this target audience, a significant strength of the book is its generally balanced discussion of the health risks posed by environmental agents. For example, although Burdon emphasizes concerns about the potential risk of cancer from environmental radiation and chemicals, this is balanced by a discussion of the endogenous risk from chemical radical attack on DNA during normal cellular metabolism. Are there significant omissions or ambiguities in the book? I will address two that I thought to be important.

First, given that risks of cancer from genotoxic environmental agents generally involve low levels of exposure, I was surprised to find so little discussion of a long-standing debate concerning dose and response. There is a school of thought (E. J. Calabrese and L. A. Baldwin *Nature* **421**, 691–692; 2003) that, at these low doses, there is an initial dose interval where cancer risk may be discounted (a threshold for risk) or where there may even be health benefits for the exposed individual (hormesis). In general, these proposals have not been viewed favourably by the national and international scientific bodies that consider risks from physical and chemical genotoxic agents. A discussion of the scientific components of the debate would have been most valuable.

My second point concerns one of the central arguments on the contribution of environmental factors to human cancer risk. Burdon emphasizes proposals that environmental factors can account for 80–90% of human cancers, and that the increasing cancer burden in the population “parallels the gradual industrialisation of the world and the widespread introduction of new synthetic chemicals into the workplace”. Although he refers to difficulties in quantifying the potential impact of environmental carcinogens and the importance of other environmental influences such as diet, I sense a degree of ambiguity in his approach. Consequently, lay readers could be forgiven for interpreting some parts of the text as reflecting scientific support for the view that involuntary exposure to environmental carcinogens is a major determining factor in human cancer risk. In my opinion, it is unlikely to be so.

In particular, although there is qualified acceptance of the view that heritable factors may account for 10–20% of cancer in Western populations, this does not imply that the remaining 80–90% of cases can be attributed

Exhibition

Unnatural causes

Successive technological revolutions have had an impact on both the natural world and society, radically changing their relationship. The Tenth International Biennale of Photography, entitled *In Nature* and showing in Turin, Italy, until 12 October, reflects the individual relationships of more than 30 artists with nature in an era of scientific dominance.

Death is a frequent theme in this powerful exhibition. For example, British documentary photographer Clive Landon recorded the horror of the livestock slaughter and burning during the 2001 outbreak of foot-and-mouth disease. One of his photographs recalls the pastoral ideal of a John Constable painting — but a closer look reveals that all the cows are dead.

Kiev artist Ilya Chichkan responds to the darkness of modern Ukrainian history. Ten years ago he ‘borrowed’ the deformed fetuses from mothers living in Kiev during and after the Chernobyl nuclear accident. The fetuses were preserved in formalin at the University of Kiev's medical school. Chichkan dressed them in jewels, like the sleeping princes of Ukrainian legend, and photographed them, as shown here, in the serene poses of a normal sleeping child, imposing an anachronistic dignity.

Alison Abbott



I. CHICHKAN

to environmental carcinogens. Although exposure to carcinogens, especially tobacco smoke, is certainly important, the non-genetic component of cancer incidence in a given population incorporates a broad range of interacting elements that may be expected to vary over time with social and economic change. These include population structure and lifestyle-related factors such as diet, reproduction and certain forms of infection (R. N. Hoover *N. Engl. J. Med.* **343**, 135–136; 2000 and www.iarc.fr/WCR).

Given their importance to a major theme of the book, these issues could have been discussed with greater clarity.

Overall, however, I found *The Suffering Gene* to be a readable and reasonably balanced introduction to DNA-related environmental issues. The discussion is a little speculative or lacking in depth in places, but given the range of issues considered, this is probably inevitable.

Roger Cox is at the National Radiological Protection Board, Chilton, Didcot OX11 0RQ, UK.

Reissued classics

The Life of Erasmus Darwin

by Charles Darwin (edited by Desmond King-Hele)
Cambridge University Press, £17.50
First unabridged edition.

Feynman Lectures on Gravitation

by Richard P. Feynman, Fernando B. Moringo & William G. Wagner
Westview, £24.99

Philosophers at War: The Quarrel Between Newton and Leibniz

by A. Rupert Hall
Cambridge University Press, £22.95

Gehennical Fire: The Lives of George Starkey, an American Alchemist in the Scientific Revolution

by William R. Newman
University of Chicago Press, \$27.50, £19.50
“Newman shows us how studying an obscure and ambiguous figure can bring the science of a period to life.” David Knight *Nature* **373**, 669 (1995).

The Making of Memory: From Molecules to Mind

by Steven Rose
Vintage, £8.99
A revised edition.

The whole tooth

What was your first experiment as a child?

I grew a colony of bacteria. I remember my father's embarrassed expression as he asked fruitlessly for agar-agar in all the pharmacies of the district.

Whose graduate student would you most like to have been?

Albert Dahlberg at the anthropology department of the University of Chicago in the early 1970s.

What scientific paper changed your career path?

Loring Brace's papers on dental size reduction in hominid evolution. Papers such as 'Post-Pleistocene changes in the human dentition' (*Am. J. Phys. Anthropol.* **34**, 191–203; 1971) were a tentative application of the 'probable mutation effect' model published in 1964 in the *American Naturalist*. Coarsely, large teeth means hard food, small teeth means soft food. Just three days before our wedding, my fiancée said that I seemed more interested in these papers than our marriage. She left, and we never saw each other again.

What book has been most influential in your scientific career?

Paleopathology at the Origins of Agriculture by M. N. Cohen and G. J. Armelagos (Academic Press, Orlando, 1984) elegantly showed 20 years ago that even the soundest ideas about recent human history are based on prejudice.

What's your favourite conference destination, and why?

Field work provides the ideal conference destination. Once back at the camp after hours spent looking for fossils and sites, sampling rocks and sediments, and excavating unusual objects, people meet to organize and summarize the results, ask questions, offer critical remarks and suggest more productive search strategies for the next day — and to decide who is in charge of the kitchen.

What book is currently on your bedside table?

Africa by John Reader, a cocktail of history, nature, adventure and politics.

What music heads the playlist in your car or lab?

Almost any opera by Rossini.

The job of captain on the Starship Enterprise in Star Trek has become vacant. Nominate any real person, living or dead, for the post.

Easy — Tim White, a palaeoanthropologist at the University of California, Berkeley.

Assuming the dead can be raised and/or time travel exists, with whom from the world outside science would you most like to have dinner?

One of the members of the early human community fossilized some 300,000–400,000 years ago in a natural pit at Atapuerca, Spain.

Where and when would you most like to have lived or worked?

Tierra del Fuego, in the second half of the nineteenth century.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

A few months after my arrival in Poitiers, I realized that two students on the same bus were joking about my French pronunciation during lectures. I strategically plunged into an article randomly extracted from my bag.

What one thing would you rescue from your burning laboratory?

Probably the collection of X-ray films of fossil specimens.

What do you most dislike about having research published?

Nothing. I'm convinced that to publish scientific results is simply the expected, proper conduct for professionals.

The Internet is the bane of scientists' lives because...

...you can be much more easily contradicted when you state to administrators that the equipment you have asked for is certainly the best value for money on the market.

What do you do to relax?

Prepare spicy sauces for dinner.

What previously under-recognized sport or pastime should be included in the Olympic Games?

Chess.

What single discovery, invention or innovation would most improve your life?

A magic pill, specific for each language, to be taken when travelling, effective for a few days, with no side effects or contra-indications.

Do you have a burning ambition to do or learn something of no practical or immediate value?

To walk across a continental interior — preferably Eurasia moving eastwards — until I reach the sea.

Under what conditions do you have your greatest and most inspired ideas?

I have a notebook labelled 'Memo & To Do'. Each morning I spend a few minutes reading and updating it. This activity represents a major source of frustration. But the unintentional visual juxtaposition of otherwise



Roberto Macchiarelli

Roberto Macchiarelli is a palaeoanthropologist. Until recently at the National Museum of Prehistory and Ethnography in Rome, he is currently professor of human palaeontology at the University of Poitiers, France.

independent listed points sometimes generates unexpected and intriguing results.

What's the one thing about science that you wish the public understood better?

That doubts, additional questions, argument and criticism contribute to the strength, not the weakness, of scientific thought.

Which field in science (apart from your own) deserves more funding, and why?

Research into the deep oceans. It may yet reveal crucial information for the life of us surface residents.

What's the most interesting thing in your fridge?

As I'm in France, capers from Puglia, jealously preserved and consumed, in very limited quantity and only with very close friends.

Which actor would best portray you in a film of your life story?

Dustin Hoffman.

What one thing would you change about Nature?

Nature & Culture might be a more appropriate title for the journal in view of its actual content.

You've just been told (in confidence) that the world will end tomorrow. What do you do next?

I would suggest to my family and friends that we take a full day of vacation together. No school, no job ... but some shopping, a visit to an exhibition, the cinema, a restaurant (probably cous-cous or paella), a drink in a pub, and then back home, with some music before bed. Then, good night... ■

Professional secrets

Craig R. Roy

Looking inside the compartments of certain immune cells — professional antigen-presenting cells — has revealed how the immune system can trigger a cell-killing response to extracellular pathogens.

Our immune system provides effective protection against most pathogens. It can mount a highly specific 'adaptive' response that distinguishes between different sorts of pathogen and activates the appropriate mechanisms to eliminate them. Writing in this issue, Amigorena and colleagues¹ and Desjardins and colleagues² suggest how an adaptive immune response typically reserved for pathogens that replicate inside immune cells can also be generated against pathogens that exist outside these cells.

Macrophages and dendritic cells are specialized cells that initiate adaptive immune responses. They are sometimes called professional antigen-presenting cells (APCs) because one of their primary duties is to degrade microbial proteins and display the resulting peptide fragments, or antigens, on their surface. The kind of immune response that is then produced depends on how the antigens are displayed by the APCs.

In general, peptide antigens from pathogens replicating inside APCs (endogenous antigens) are transported into a cellular compartment, the endoplasmic reticulum (ER), and bind to a protein complex called MHC class I. From the ER, the MHC class I–antigen complex is shuttled to the cell surface, where it stimulates antigen-specific cytotoxic T lymphocytes (CTLs) — killer cells that seek out and eliminate infected cells (Fig. 1).

Antigens from extracellular pathogens are usually processed differently from endogenous antigens, and with different results. Extracellular pathogens are engulfed by APCs and degraded in phagosomes — membrane-bound cellular compartments. Peptide fragments produced in phagosomes are not transported to the ER. Instead, they bind directly to a different set of MHC molecules, MHC class II, inside the phagosome and then return to the cell surface, where they stimulate a subset of helper T cells (Fig. 1). Unlike the CTLs, helper T cells do not kill other cells, but instead produce chemical signals. These trigger the production of antigen-specific antibodies and elicit an inflammatory response that eliminates the microbial pathogens.

So the rules governing antigen presentation seem to be simple — endogenous antigens are presented by MHC class I molecules to stimulate a cell-killing response, and exogenous antigens are presented by MHC class II molecules to stimulate a helper

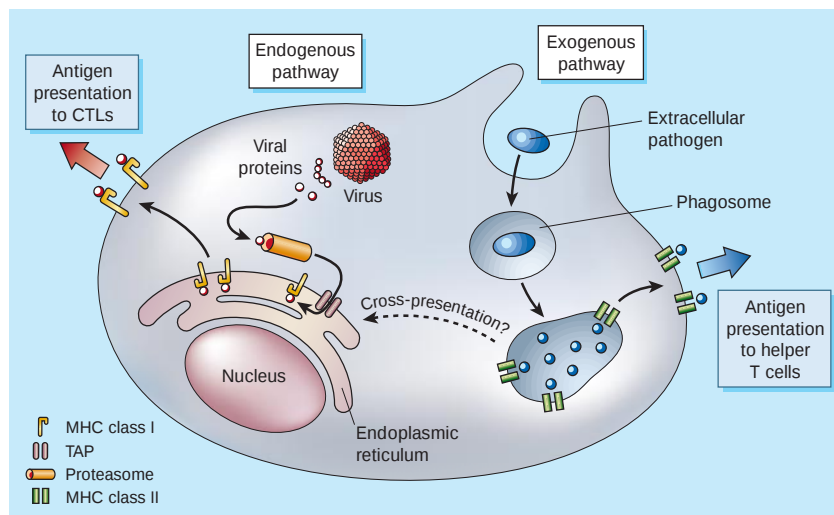


Figure 1 Professional antigen-presenting cells process intracellular and extracellular pathogens differently. In the endogenous pathway, proteins from intracellular pathogens, such as viruses, are degraded by the proteasome and the resulting peptides are shuttled into the endoplasmic reticulum (ER) by TAP proteins. These peptides are loaded onto MHC class I molecules and the complex is delivered to the cell surface, where it stimulates cytotoxic T lymphocytes (CTLs) that kill the infected cells. In contrast, extracellular pathogens are engulfed by phagosomes (exogenous pathway). Inside the phagosome, the pathogen-derived peptides are loaded directly onto MHC class II molecules, which activate helper T cells that stimulate the production of antibodies. But some peptides from extracellular antigens can also be 'presented' on MHC class I molecules. How this cross-presentation occurs has now been explained^{1,2}. It seems that by fusing with the ER, the phagosome gains the machinery necessary to load peptides onto MHC class I molecules (see Fig. 4g on page 405).

response. But in reality, the situation is not so straightforward. Professional APCs can break these rules and present exogenous antigens on MHC class I molecules³, in a process known as cross-presentation. The cross-presentation pathway seems to be required for CTL responses to certain exogenous antigens^{4,5}. For example, some viruses do not infect professional APCs, but in order to generate a CTL response that will eliminate virus-infected cells, the APC must present exogenous viral proteins on MHC class I molecules. Cross-presentation might also be important for triggering CTL responses to some cancer cells.

So how do exogenous antigens cross over to the MHC class I pathway? Amigorena and colleagues had previously shown⁶ that exogenous proteins could escape from phagosomes and enter the cytosol (the intracellular space). So it was thought that cross-presentation might simply involve the transport of exogenous proteins from the phagosome into the cytosol, where they

could join the pathway used to process endogenous antigens. This view was supported by data^{7,8} showing that cross-presentation depends on two components of the MHC class I pathway — the proteasome complex and the 'transporter associated with antigen presentation' (TAP). The proteasome complex degrades proteins in the cytosol and TAP shuttles the resulting peptide antigens into the ER for loading onto MHC class I molecules. Furthermore, cross-presentation is sensitive to the drug brefeldin A (refs 7, 8), which disrupts the transport of the MHC class I–antigen complex from the ER to the cell surface⁹.

Desjardins and colleagues, meanwhile, had provided a clue to how exogenous antigens might escape from the phagosome^{10,11}. In an effort to understand how phagosomes form and mature, these authors had catalogued the proteins contained on phagosomes that had been purified from macrophages. Surprisingly, they detected proteins that were thought to reside only in the ER¹⁰. Further analysis

revealed that phagosomes fused with the ER during, or shortly after, pathogen engulfment at the cell surface¹¹. This was an important finding because a protein-translocation channel called the Sec61 complex is embedded in the ER membrane. Sec61 both imports newly synthesized proteins into the ER and exports proteins from it, targeting the proteins for degradation by the proteasome^{12,13}. So it was predicted that the Sec61 complex might be delivered to the phagosome after fusion with the ER. If true, it could be the missing link in the cross-presentation pathway — the transporter responsible for exporting exogenous proteins from the phagosome into the cytosol.

The two groups^{1,2} now provide evidence to support this hypothesis. Looking at phagosomes from dendritic cells¹ and macrophages², respectively, they showed that Sec61 is indeed present on the phagosome membrane. And both groups found that a fluorescently tagged version of the protein ovalbumin could be exported from the phagosome into the cytosol. But did protein export involve Sec61? To find out, Desjardins and colleagues looked at the export of the cholera toxin A1 subunit (CTA1). This protein moves from outside the cells to the ER and is then exported from the ER into the cytoplasm through the Sec61 channel^{14,15}. The authors observed that CTA1 could also be exported from the phagosome into the cytoplasm, which strongly suggests that, after Sec61 is delivered to the phagosome, it can still export proteins.

What happens to proteins once they have been exported from the phagosome? First, they must be turned into peptide antigens by the proteasome. Desjardins and colleagues showed that proteasomes are associated with the cytosolic side of the phagosome membrane. And both groups showed that TAP and the MHC class I molecules are delivered to the phagosomes and remain active. Their further analyses revealed that the peptide antigens are loaded onto MHC class I molecules inside the phagosome.

So it seems that after the exogenous proteins are exported from the phagosome through the Sec61 channel, they are degraded by the proteasome and the resulting peptide antigens are shuttled back into the phagosome by TAP. There, they are loaded onto MHC class I molecules on the inside of the phagosome membrane. The results confirm observations¹⁶ that the phagosome is a fully competent antigen-processing compartment for the MHC class I pathway.

The new studies provide strong support for a model of cross-presentation in which ER-phagosome fusion occurs. But they do not show that such fusion is necessary. Arguments have also been made that cross-presentation occurs only in cultured cells and might not be relevant in animals¹⁷. Now that a molecular mechanism governing cross-presentation has been proposed,

experiments to test the importance of this immunological process are certain to follow. But the debate on cross-presentation continues — to resolve it, the professionals will have to reveal all of their secrets. ■

Craig R. Roy is in the Section of Microbial Pathogenesis, Yale University School of Medicine, New Haven, Connecticut 06536, USA.

e-mail: craig.roy@yale.edu

1. Guermonprez, P. *et al.* *Nature* **425**, 397–402 (2003).
2. Houde, M. *et al.* *Nature* **425**, 402–406 (2003).
3. Bevan, M. J. *J. Exp. Med.* **143**, 1283–1288 (1976).
4. den Haan, J. M. & Bevan, M. J. *Curr. Opin. Immunol.* **13**, 437–441 (2001).
5. Sigal, L. J., Crotty, S., Andino, R. & Rock, K. L.

- Nature* **398**, 77–80 (1999).
6. Rodriguez, A., Regnault, A., Kleijmeer, M., Ricciardi-Castagnoli, P. & Amigorena, S. *Nature Cell Biol.* **1**, 362–368 (1999).
7. Brossart, P. & Bevan, M. J. *Blood* **90**, 1594–1599 (1997).
8. Kovacsics-Bankowski, M. & Rock, K. L. *Science* **267**, 243–246 (1995).
9. Nuchtern, J. G. *et al.* *Nature* **339**, 223–226 (1989).
10. Garin, J. *et al.* *J. Cell Biol.* **152**, 165–180 (2001).
11. Gagnon, E. *et al.* *Cell* **110**, 119–131 (2002).
12. Koopmann, J. O. *et al.* *Immunity* **13**, 117–127 (2000).
13. Romisch, K. J. *Cell Sci.* **112**, 4185–4191 (1999).
14. Roy, C. R. *Trends Microbiol.* **10**, 418–424 (2002).
15. Schmitz, A., Herrgen, H., Winkler, A. & Herzog, V. J. *Cell Biol.* **148**, 1203–1212 (2000).
16. Ramachandra, L., Song, R. & Harding, C. V. J. *Immunol.* **162**, 3263–3272 (1999).
17. Zinkernagel, R. M. *Eur. J. Immunol.* **32**, 2385–2392 (2002).

Plasma physics

Cosmic waves in the lab

Rod Boswell

An Alfvén-wave maser, a feature of atmospheric and astrophysical science, has been created in a laboratory, and opens the way for further Earth-bound investigations of cosmic phenomena.

In general, to be an observational astronomer is to be a spectroscopist, unravelling the workings of the cosmos through the varying wavelengths of radiation detected. Observations that began at optical wavelengths now extend to wavelengths ranging from hundreds of kilometres down to tens of nanometres, and satellites have also proved immensely useful in refining our image of the Universe. But, despite the assiduous measurements, the radiation is often produced, particularly at radio and X-ray wavelengths, by very complex processes that we struggle to understand. Nevertheless, a positive step has now been taken: in *Physical Review Letters*, J. E. Maggs and G. J. Morales report the resonant amplification of a typical astrophysical wave in their laboratory — an Alfvén-wave maser

(*Phys. Rev. Lett.* **91**, 035004; 2003).

For some years, the aim of these authors has been a laboratory study of magneto-hydrodynamic turbulence — that is, turbulence in the interactions between a plasma and a magnetic field. The first step on the way is to create the basic element of the phenomenon, an Alfvén wave. Postulated in 1942 by Hannes Alfvén (*Nature* **150**, 405–406; 1942), an Alfvén wave is a low-frequency electromagnetic wave that can be generated in magnetized plasmas throughout the Universe. These waves are thought to be intimately involved in diverse phenomena, such as the precipitation of electrons through the atmosphere in the auroral regions that are connected with the dazzling northern and southern lights — and that can take out power distribution systems. Alfvén waves are also implicated in the heating



Figure 1 Unique facility — the Large Plasma Device at the University of California, Los Angeles. In a helium plasma inside this 19-metre-long column of magnets, Maggs and Morales have created an Alfvén-wave maser.

of electrons in the solar corona and in the dynamics of solar flares.

In many situations, an Alfvén wave might not propagate freely but instead be trapped between reflective surfaces, for example between the Sun's photosphere and corona, in the intervening layer known as the chromosphere. This is analogous to light waves trapped between the mirrors of a resonant cavity of a laser. Laser action at microwave frequencies — a 'maser' — is a known astrophysical phenomenon. So Maggs and Morales set out to demonstrate a laboratory-based maser for Alfvén waves.

In general, laboratory experiments attempting to mimic the behaviour of the aurora, the solar chromosphere or more exotic cosmic objects have met with scepticism, if not outright hostility, in the astrophysical community. Ever since William Gilbert, in about 1600, used a magnetized sphere — a terrella — to explain the magnetic compass to Queen Elizabeth I, many brave attempts have been made in terrestrial laboratories to pin down exotic cosmic phenomena. Designing a terrestrial experiment to study Alfvén waves is certainly not a simple exercise — several conflicting scalings need to be satisfied if the sceptics are to be convinced. For example, it is extremely difficult to make a laboratory experiment long enough and wide enough to allow the wave to propagate in a natural way and not have it cramped up and its motion constricted. Also, the laboratory plasma needs to be created from a background gas, which will collide with the plasma and damp out waves. This really does not exist in the vacuum of space.

Maggs and Morales' experiment was performed using the Large Plasma Device (at the Basic Plasma Science Facility at the University of California, Los Angeles), a unique facility with 90 magnetic-field coils surrounding a column of helium plasma 19 m long (Fig. 1). At one end of the column, a pulsed voltage draws electrons out of a thermionic cathode towards an anode made of copper mesh, located 55 cm away. The accelerated electrons drift into the 19-m vacuum chamber and strike neutral, low-pressure helium gas, generating a plasma that is more than 50% ionized.

The resonant cavity is formed by the space between the cathode and the semi-transparent anode: inside, Alfvén waves are spontaneously amplified, but they seep through the anode into the main plasma region, in which they can be detected. The magnetic field, the plasma density and the degree of ionization need to be manipulated so that the ion cyclotron frequency (the frequency with which the positively charged ions rotate around the magnetic field) is much greater than the ion neutral collision frequency (the rate at which plasma collides with the background gas). This means that the Alfvén waves are not damped in the

plasma column. The Large Plasma Device is one of the very few experimental devices in the world in which such a happy combination of operational parameters can be achieved.

Using small magnetic loop antennae inserted into the plasma column, Maggs and Morales measured a broadband noise spectrum with a peak at 60% of the ion cyclotron frequency, and an upper limit of 70% of that frequency defined by the length of their resonant cavity. This noise was identified as a type of Alfvén wave called a shear wave. The *pièce de résistance* of the experiment was their observation of spontaneous amplification of the resonant mode (at 60% of the ion cyclotron frequency) for selected values of the confining magnetic field and plasma current. Maggs and Morales describe the phenomenon as "spectacular flares of

extremely coherent signals [that] develop at low magnetic fields and extremely high plasma currents" — a little reminiscent of sci-fi writer E. E. 'Doc' Smith perhaps, but nevertheless explosively descriptive of the growth and subsequent decay of their Alfvén maser.

Maggs and Morales' experiment demonstrates the great potential for carrying out experiments of cosmic importance in the laboratory. It will no doubt be a boon for the many theoreticians wishing to test their theories of magnetohydrodynamic turbulence and of Alfvén-wave-induced heating of stellar atmospheres. ■

Rod Boswell is in the Plasma Research Laboratory, Research School of Physical Sciences and Engineering, Australian National University, Canberra, ACT 0200, Australia.
e-mail: rod.boswell@anu.edu.au

Stem cells

To be and not to be

Haifan Lin

It has long been proposed that stem cells function by dividing to generate an identical daughter cell and a cell that becomes more specialized. New work illustrates such asymmetric division and its molecular basis.

Stem cells have the unique ability to perpetuate themselves while continually replenishing tissues throughout the life of an organism. This ability has long been attributed to a distinctive asymmetry in their division, such that when a stem cell divides, it gives rise to both an exact copy of itself and a new type of cell that will differentiate into mature cells of the tissue. This asymmetry hypothesis — reflecting a 'to be and not to be' fate decision — can be traced back to the embryologist E. B. Wilson and his contemporaries more than a century ago. But direct evidence for asymmetric division has so far been reported in only a few well-defined stem-cell systems. In these cases, the molecular mechanisms that regulate asymmetric division have also been explored. Work by Yamashita and colleagues¹, reported in *Science*, now showcases an elegant example of asymmetric division, and identifies key new molecules involved in the process.

A hefty roadblock to the study of how stem cells divide has been their elusive nature. Stem cells are rare, and physically resemble their differentiating daughter cells, so they are difficult to recognize in a tissue. It is fair to say that even today — and despite rapid progress in stem-cell research — the identity of stem cells in most tissues remains a matter of debate. (Embryonic stem cells, which are well characterized, exist before tissues form.) This 'identity crisis' has been overcome in only a few cases, such as neuroblasts and germline stem cells in the fruitfly *Drosophila*,

and ventricular-zone progenitor cells in the brains of mammalian embryos^{2,3}.

In flies, a neuroblast divides asymmetrically (Fig. 1a, overleaf) to renew itself and to produce a smaller ganglion mother cell², which goes on to generate specific nerve cells. In fly ovaries, germline stem cells are always in contact with a subset of signalling cells called cap cells³. These stem cells also divide asymmetrically, producing a stem-cell daughter that retains contact with the cap cells, and a differentiating daughter that is displaced one cell away⁴ (Fig. 1b). Such oriented asymmetric division has also been demonstrated by electron microscopy for germline stem cells in the fly testis⁵, where the signalling cells are called hub cells (Fig. 1c). Yamashita *et al.*¹ have used immunofluorescence microscopy to observe a large number of stem-cell divisions (more than 500) in fly testes, nicely confirming the orientation and asymmetry of division.

The key question here is: what tells one daughter cell to be a stem cell and the other not to be? For fly neuroblasts, the solution seems to depend solely on the cell itself². The neuroblasts are derived from embryonic epithelial-type cells, and inherit their polarity, with one end being 'apical' and the other 'basal'. This allows molecules that determine cell fate to be segregated along the apical-basal axis. The mitotic spindle (the structure that separates chromosomes during division) is also oriented along this axis, such that the plane of division is perpendicular to the

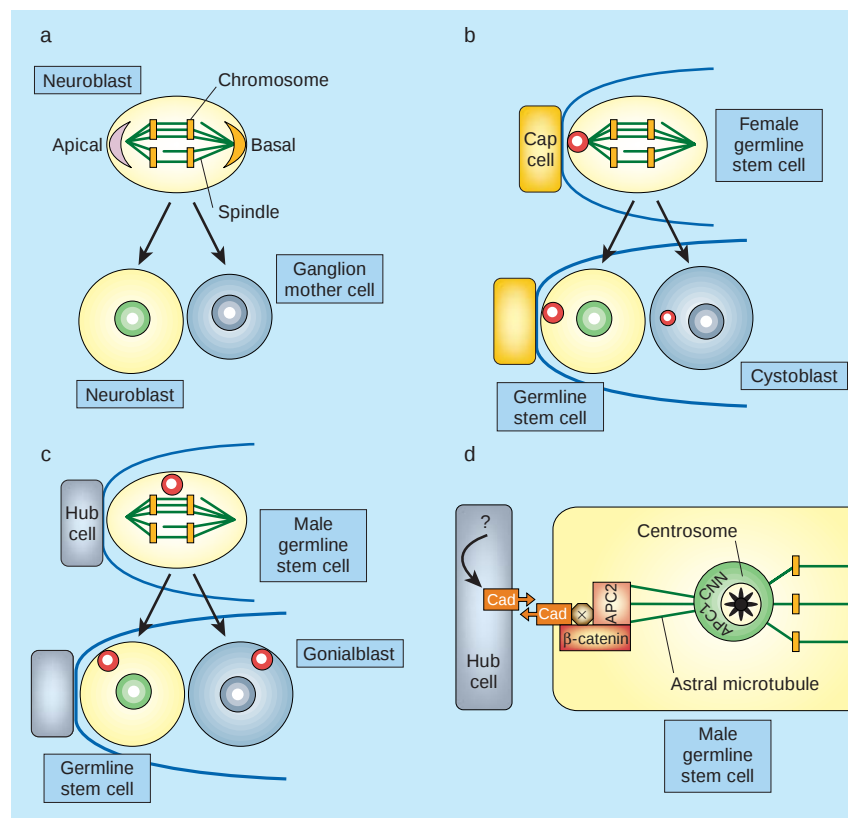


Figure 1 Stem-cell decisions. a, A dividing fly neuroblast gives rise to a neuroblast and a smaller ganglion mother cell. The protein complexes (crescents) at the apical and basal ends of the dividing neuroblast are different, ensuring the asymmetry of division. b, A germline stem cell in the fly ovary, associated with a cap cell, produces a germline stem cell and a cystoblast. The spectrosome (red) anchors one pole of the mitotic spindle to the cap cell to define the orientation of division. c, A germline stem cell in the fly testis, associated with hub cells, produces a daughter germline stem cell and a gonialblast. Yamashita *et al.*¹ have found that the spectrosome is not involved in this division. d, Yamashita *et al.* propose that the orientation of division in male germline stem cells is determined by the proteins DE-cadherin (Cad), β -catenin, centrosomin (CNN), adenomatous polyposis coli (APC) and an unidentified component, X.

axis. This means that one daughter cell inherits the apical molecules and remains a neuroblast; the other inherits the basal components and becomes a ganglion mother cell.

But this self-sufficient mechanism does not seem to be enough for most other stem cells. For instance, in fly ovaries, stem cells must remain within the stem-cell 'niche' — in contact with cap cells — to maintain their fate³ (Fig. 1b). Here, the oriented asymmetric division ensures that, following each division, only one daughter cell will have contact with cap cells and so remain a stem cell. A similar situation occurs for male germline stem cells^{1,5}.

So these three types of stem cell share one feature. The orientation of division — in turn determined by the orientation of the spindle — is key to ensuring an asymmetry between the daughter cells, either in their exposure to a niche signal or in their inheritance of fate-regulating molecules. Then what determines spindle orientation? For female germline stem cells, this requires a spherical organelle called the spectrosome⁴, which at one spindle

pole anchors filamentous structures called astral microtubules to the cap cells. But Yamashita *et al.*¹ show that the spectrosome is often not associated with the spindle pole in male germline stem cells, and thus may not orient the asymmetric division. So the function of the spectrosome might depend on the cellular milieu.

What, then, determines spindle orientation in testis stem cells? Previous work on flies had begun to reveal the importance, in other cell divisions, of cell-to-cell junctions called adherens junctions, and of molecules involved in linking the spindle to these junctions. One such molecule in neuroblasts is centrosomin⁶, a component of the centrosome. (Dividing cells have two centrosomes, one at each pole; they produce the spindle and astral microtubules.) In the ovary, the proteins DE-cadherin and β -catenin are important for anchoring germline stem cells to cap cells⁷. In embryonic epithelial cells, disrupting adherens junctions causes misorientation of the spindle⁸, and the adenomatous polyposis coli 2 protein (APC2), a

binding partner for β -catenin, binds to adherens junctions⁸. Finally, APC2 and β -catenin anchor spindles to the periphery in early fruitfly embryos⁹.

Yamashita *et al.*¹ show that such a linkage may also define the orientation of division of male germline stem cells. They observed that DE-cadherin, β -catenin and APC2 accumulate at the interface between the stem cell and hub cells, where one pole of the stem-cell spindle is pinned (Fig. 1d). In flies with mutations in APC2, spindles become mispositioned, misoriented or detached from the interface. APC1, meanwhile, localizes to centrosomes just before cell division, and in flies with mutations in APC1, spindle orientation and centrosome position are also somewhat perturbed. So the authors suggest that the adherens junctions, via β -catenin and APC2, anchor the astral microtubules of one of the stem cell's centrosomes. This in turn pins the spindle pole in place. Meanwhile, centrosomin and APC1 contribute to the centrosome's function in spindle-pole organization.

Interestingly, in budding yeast, which undergo stem-cell-like divisions, an APC-like protein called Kar9 has a similar role in capturing microtubules. All of these studies suggest that the spindle-anchorage mechanism described above has been conserved for asymmetric stem-cell division during evolution.

Of course, the stereotypical asymmetric division described here is probably not the only way for stem cells to self-renew. For example, certain mammalian stem cells appear to divide symmetrically¹⁰. Consequently, both daughter cells may sometimes remain as stem cells, because they are both exposed to the niche signal or receive appropriate fate determinants; sometimes both may differentiate. For these stem cells, self-renewal is based on a probabilistic mechanism. But even though division is oriented at random, the nature of the niche signalling and the localized determinants should be the same as those in the asymmetrically dividing stem cells. So stereotypical stem cells, with their predictable pattern of division, can serve as simplified models of their symmetrically dividing counterparts. The findings of Yamashita *et al.*¹ thus represent exciting progress in understanding both types of stem cell.

Haifan Lin is in the Department of Cell Biology, Duke University Medical School, Durham, North Carolina 27710, USA.
e-mail: hlin@duke.edu

1. Yamashita, Y. M., Jones, D. L. & Fuller, M. T. *Science* **301**, 1547–1550 (2003).
2. Jan, Y.-N. & Jan, L. Y. *Nature Rev. Neurosci.* **2**, 772–779 (2001).
3. Lin, H. *Nature Rev. Genet.* **3**, 931–940 (2002).
4. Deng, W. & Lin, H. *Dev. Biol.* **189**, 79–94 (1997).
5. Hardy, R. W., Tokuyasu, K. T., Lindsley, D. L. & Garavito, M. *J. Ultrastruct. Res.* **69**, 180–190 (1979).
6. Megraw, T. L., Kao, L. R. & Kaufman, T. C. *Curr. Biol.* **11**, 116–120 (2001).
7. Song, X., Zhu, C. H., Doan, C. & Xie, T. *Science* **296**, 1855–1857 (2002).
8. McCartney, B. M. *et al.* *Nature Cell Biol.* **3**, 933–938 (2001).
9. Lu, B. *et al.* *Nature* **409**, 522–525 (2001).
10. Watt, F. M. & Hogan, B. L. M. *Science* **287**, 1427–1430 (2000).



100 YEARS AGO

Aéronautics. How little we appear to have advanced beyond where we were fifty years ago, when on September 24, 1852, that eminent French engineer, Henri Giffard, succeeded during an experimental ascent in Paris in driving a balloon against the wind for a very short distance, although on October 19, 1901, M. Santos Dumont was successful in navigating his balloon from St. Cloud round the Eiffel Tower in Paris and back to the spot where he had started only half an hour previously. Many have been engaged in this so far unsolved problem of aerial navigation, but there is one of whom we seldom hear. I will quote what Dr. Janssen said in his Presidential Address to the International Aéronautic Congress... regarding Mr. Langley, Correspondent of the Institute of France and Secretary of the Smithsonian Institution at Washington. "Independently of the fine and profound researches of this investigator upon the resistance of air, Mr. Langley has constructed an aéroplane which has progressed and has sustained itself during a time notably longer than any of the apparatus previously constructed." In the last report of the Smithsonian Institution, that for 1901, it is stated that this steel flying-machine had... a weight of 30 lb., developed $1\frac{1}{2}$ horse-power, and repeatedly flew from one-half a mile to three-quarters of a mile.

[The writer's bet on the 'aéroplane' having a future was soon borne out. Before the year was out, the Wright brothers had made aviation history with their flight at Kitty Hawk, North Carolina.]

From *Nature* 24 September 1903.

50 YEARS AGO

Traité de paléontologie. Publié sous la direction de Prof. Jean Piveteau. Space forbids a review of individual chapters, but many general points invite criticism. From Prof. Piveteau's introduction it would appear that nothing significant in palaeontology has ever been done outside France (with one Belgian exception). This attitude has fortunately not affected the contributors. The chapters on general principles display that characteristically French blend of authority with enthusiasm which disarms criticism... these two volumes will be found more valuable on general questions than as taxonomic handbooks. The best things have the true Gallic brilliance and penetration and will endure. The weaker parts show that so formidable a task as this is not to be lightly undertaken.

From *Nature* 26 September 1953.

Synthetic chemistry

Ship reverses out of bottle

Avelino Corma

The industrial application of zeolites is limited by the cost of certain organic materials that are needed to make them, but which are destroyed in the process. A clever technique offers a solution.

Zeolites are perhaps best known as molecular sieves. These porous structures occur naturally, but are more familiar as products of the ingenuity of chemists — as, for instance, components of water softeners. Their regular pore structure and high pore volumes allow them accurately to discriminate between, and organize, molecules. This characteristic means that synthetic zeolites have found wide application as 'nanoreactors' and selective membranes, and even as components of electronic devices¹.

Scientists in both the academic and industrial worlds are working hard to find ways to make new zeolitic materials with more complex pore topologies, and to improve the existing synthetic techniques. One fruit of their labours is to be found on page 385 of this issue², where Lee *et al.* describe a promising method for recovering and recycling the organic 'structure-directing agents' (SDAs). These molecules are used to control pore formation in zeolite synthesis, but they tend to be expensive and typically are destroyed during the production process.

Zeolites are crystalline aluminosilicates, made of interlocked tetrahedrons of SiO_4 and AlO_4 in which each oxygen is shared between adjacent tetrahedra. New structures can be created through the directing effect of some of the framework cations, which favours the presence of certain subunits in the final zeolite structure^{3–6}, but more especially by using new SDAs^{7,8}. In the latter case, the self-assembly of the crystalline inorganic zeolite is dictated by a complex process that involves numerous weak interactions with the SDA concerned (water-soluble quaternary ammonium ions, amines and crown ethers are examples often used). The SDA determines the pore topology by filling the pore space and maintaining a close geometrical correspondence with the inorganic framework. The organic components are then burned off, leaving a zeolite with a specific porosity.

In many cases, the creation of a novel zeolite structure requires the preparation of complex and costly SDAs, but their destruction during zeolite synthesis is economically prohibitive. Furthermore, the high temperatures (500 °C or more) and the water formed during SDA removal may break some of the structural Si–O or Al–O bonds. This, in turn, can destroy potentially catalytic active sites or reduce the crystallinity of the final zeolite. A way of making zeolites at low tempera-

tures, and allowing the SDA to be recovered and recycled, would be highly desirable.

Lee *et al.*² have devised just such a method. Because their procedure does not require a high-temperature processing step, it does not have the drawbacks stemming from that part of the conventional procedure. Conceptually, the SDA concerned has several virtues: it is stable at the high pH required for the synthesis; it meets the conditions of rigidity, size and polarity required⁷; and it can be easily disassembled inside the finished zeolite's pores and removed.

Lee *et al.* have proved their concept by synthesizing the zeolite ZSM-5 using a type of molecule known as a cyclic ketal. This was prepared by reacting a cyclic ketone with ethylene glycol — a reversible reaction, as shown schematically in the paper (see page 386). The ketal molecules remain stable during the zeolite synthesis and end up occluded within the pores of the final crystalline structure. Their size means that they cannot pass out of the pores and under usual practice they would be burned off and lost. But selection of a cyclic ketal as the SDA is clever, because these molecules can easily be broken up into the original cyclic ketone and ethylene glycol by lowering the pH under mild liquid or gas conditions. The ethylene glycol can simply be flushed out of the zeolite structure, and the ketone, which is smaller than the cyclic ketal, can be removed intact from the pores by an ion exchange with NaOH and NaCl at 100 °C. This process of disassembly is the reverse of the ship-in-a-bottle technique, by which organic 'guests' with useful properties are introduced into the pores of zeolites.

The authors take the reaction full circle: from cyclic ketone and ethylene glycol, to the SDA cyclic ketal, to synthesis of the ZSM-5 zeolite, followed by recovery of the original glycol and ketone. Importantly, the chemical extraction and recovery of the SDA does not seem to break Si–O–Al bonds in ZSM-5, thereby preserving the acid sites and catalytic activity associated with the framework aluminium. The authors claim that their concept could be expanded to other, similar SDAs that are stable under alkaline conditions but which break up at low pH, such as acetals and orthoesters. To my mind, particularly enticing prospects centre on the large number of SDAs whose synthesis proceeds through ketone intermediates — as with those made with the Diels–Alder reaction⁹, which is one

of the most widely used reactions in organic chemistry and is often used to synthesize the SDAs needed to create large-pore or even ultra-large-pore zeolites.

The new method could perhaps also be applied to zeolite syntheses that do not necessarily involve high pH. For instance, reactions can be performed under mild acidic or neutral conditions if fluoride salts are used in the initial stages of zeolite formation. Similarly, the synthesis of microporous aluminophosphates does not require a high pH. This should expand the technique's range to more than just those SDAs that are stable under alkaline but not under acidic conditions.

Lee *et al.* show that their elegant procedure works in the laboratory. It will need to be scaled up for industrial purposes, and optimized at that level. But zeolite structures that are at present considered no more than

laboratory curiosities — because expensive, non-recyclable SDAs are required to make them — should then make their appearance in the wider world.

Avelino Corma is in the Instituto de Tecnología Química, UPV-CSIC, Universidad Politécnica de Valencia, Avenida de los Naranjos s/n, 46022 Valencia, Spain.

e-mail: acorma@itq.upv.es

1. Davis, M. E. *Nature* **417**, 813–821 (2002).
2. Lee, H., Zones, S. I. & Davis, M. E. *Nature* **425**, 385–388 (2003).
3. Annen, M. J., Davis, M. E., Higgins, J. B. & Schlenker, J. L. *J. Chem. Soc., Chem. Commun.* 1175–1176 (1991).
4. Rohrig, C. & Gies, H. *Angew. Chem. Int. Edn Engl.* **34**, 63–65 (1995).
5. Corma, A., Rey, F., Valencia, S., Jorda, J. L. & Rius, J. *Nature Mater.* **2**, 493–497 (2003).
6. Corma, A., Díaz-Cabañas, M. J., Martínez-Triguero, J., Rey, F. & Rius, J. *Nature* **418**, 514–517 (2002).
7. Lobo, R. F., Zones, S. I. & Davis, M. E. *J. Inclusion Phenom. Mol. Recognit. Chem.* **21**, 47–78 (1995).
8. Lee, S. L. *et al.* *J. Am. Chem. Soc.* **124**, 7024–7034 (2002).
9. Nakagawa, Y. *US Pat.* 5,281,407 (1994).

Cancer

The rules of attraction

Len Neckers and Yong-Sok Lee

The puzzle of how a drug that binds to a protein found in normal cells as well as cancer cells preferentially kills tumours is now solved — the target protein exists in a drug-binding complex in tumour cells.

Targeting a specific protein or a single signalling pathway that is required for the survival of tumour cells but not normal cells would seem to be a promising anticancer strategy. Unfortunately, few such unique targets exist, and it is becoming clear that inhibiting a single pathway might not

be enough to tackle cancers that result from several genetic abnormalities. Instead, attention is turning to proteins such as heat-shock protein 90 (Hsp90) that regulate many signalling pathways in cancer cells. One feature of Hsp90 has concerned investigators — although cancer cells can

produce high levels of the protein¹, it is also abundant in normal cells. This might mean that drugs targeting Hsp90 prove to be unacceptably toxic. Surprisingly, however, the first Hsp90 inhibitor to be tested in clinical trials, the drug 17-AAG, has been well tolerated by patients. On page 407 of this issue, Kamal *et al.*² provide data that begin to explain this apparent paradox. These authors report that Hsp90 found in tumour cells has a much higher affinity for 17-AAG than does Hsp90 from normal cells.

In 1962, while looking at the salivary-gland chromosomes of the fruitfly *Drosophila*, Ferruccio Ritossa noticed that certain regions of the chromosomes puffed out in response to a sudden increase in temperature³. The gene products encoded on these chromosome puffs were later isolated and termed heat-shock proteins, or Hsps. The production of Hsps accelerates in response to temperature stress, but these proteins are abundant even in unstressed cells. Hsps have been more accurately called 'molecular chaperones', because they protect other cellular proteins from becoming misshapen as a result of high temperature or other environmental insults, and certain Hsps also enable newly synthesized proteins to attain the correct conformation.

One particular chaperone, Hsp90, has been implicated in the survival of cancer cells⁴. Hsp90 regulates the function and stability of many key signalling proteins that help cancer cells to escape the inherent toxicity of their environment, to evade the effects of chemotherapy, and to protect themselves from the results of their own genetic instability. So inhibitors of Hsp90 could mount a multi-pronged assault on cancer cells that, if not lethal itself, might leave them sufficiently debilitated to allow control by chemotherapy or radiotherapy.

Despite this compelling rationale for Hsp90-directed anticancer therapy, the abundance of the protein in normal cells raised concerns that an Hsp90 inhibitor, such as 17-AAG, would be unacceptably toxic to patients. But unaccountably, pre-clinical data and the results of an initial clinical trial showed that this drug seemed to target tumour cells in preference to normal cells. Why? Kamal *et al.*² provide an answer. They find that Hsp90 derived from tumour cells binds to 17-AAG up to 100 times more tightly than does Hsp90 isolated from normal cells. Intriguingly, Hsp90 from normal cells binds the drug with nearly the same low affinity as does purified Hsp90 (refs 2, 5).

To begin to investigate the higher affinity of 17-AAG for tumour Hsp90, the authors looked at whether the chaperone interacted with other proteins inside tumour cells and normal cells. To stabilize its 'client' proteins, Hsp90 assembles with other chaperones and associated proteins to form a 'super-chaperone machine'⁶. Kamal *et al.* found that in

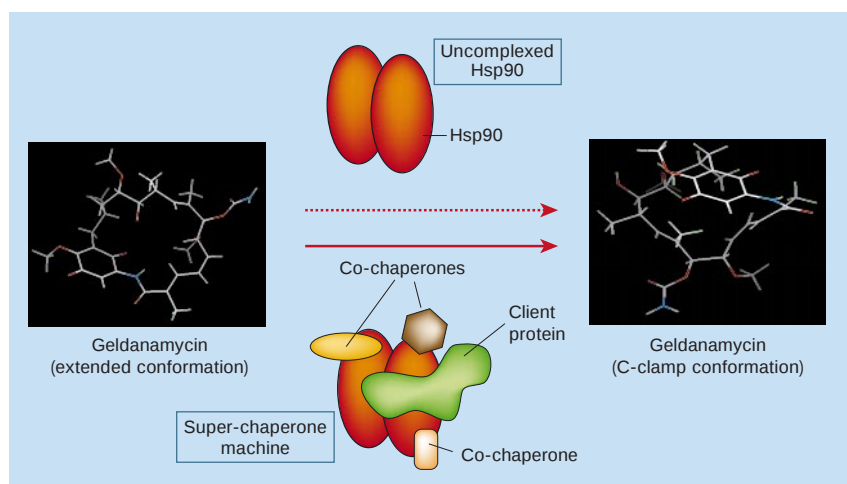


Figure 1 Catalysts of change? The Hsp90 inhibitor geldanamycin exists in two forms. In solution it exists in an extended conformation, but it must switch to a 'C-clamp' conformation before it can bind to the Hsp90 protein. The energy barrier between these two conformations is too great to allow spontaneous conversion, so the change must be catalysed. Kamal *et al.*² show that the drug 17-AAG, which is related to geldanamycin, binds more tightly to Hsp90 when the protein is part of a 'super-chaperone machine' that actively modulates the shape of 'client' proteins. It is possible that Hsp90 itself catalyses the conversion of drugs like 17-AAG — and that the super-chaperone machine might be a more efficient catalyst of this process than uncomplexed Hsp90.

tumour cells, the bulk of Hsp90 exists in such an assembly, whereas most of the Hsp90 in normal cells exists in a free form. The Hsp90 in tumour cells also had higher ATPase activity (required for its chaperone function) — a finding that supports the view that tumour Hsp90 is present in fully active chaperone complexes.

So the affinity of 17-AAG for Hsp90 seems to depend on the incorporation of the chaperone into a multi-protein machine. Strikingly, Kamal *et al.* were able to increase the weak affinity of purified Hsp90 for 17-AAG roughly 50-fold by adding components of the chaperone machine. Data on drug levels achieved in patients support these findings — intravenously administered 17-AAG is present in the circulation for many hours at concentrations far exceeding its apparent affinity for tumour Hsp90, but only transiently reaches the high concentration that would allow binding to the free form of the protein found in normal cells⁷. This might partly explain the relative lack of toxicity of 17-AAG in patients.

What makes Hsp90 better able to bind 17-AAG when the protein is part of the super-chaperone complex? At present we can only speculate. An intriguing possibility is that the chaperone complex might catalyse a conformational change in the drug. Indeed, it has been suggested^{8,9} that the structurally similar drug geldanamycin (the 'parent' of 17-AAG) must undergo a conformational change from an open, planar structure to a more compact 'C-clamp' shape before it can bind to Hsp90 (Fig. 1). The energetics of the spontaneous conversion between these two forms is highly unfavourable⁹ (Y.-S. Lee, M. G. Marcu and L. Neckers, unpublished observations). So one or more components of the super-chaperone machine might be much more efficient than free Hsp90 at catalysing a conformational change in drugs such as 17-AAG; alternatively, the complex might potentiate the ability of Hsp90 itself to do so. Both possibilities are feasible. For example, association with other chaperones has been shown to stimulate the normally weak ATPase activity of Hsp90 (ref. 10). And other components of the Hsp90 multi-chaperone complex possess an intrinsic ability to modulate client-protein conformation¹¹. Does the Hsp90 super-chaperone machine view 17-AAG as a protein in need of refolding? Further experiments will be needed to answer this question.

Meanwhile, modifications of 17-AAG and other geldanamycin derivatives are being developed that spontaneously form the C-clamp conformation in solution. It will be interesting to compare their affinity for tumour Hsp90 with that for Hsp90 from normal cells — if they do not retain preferential binding to tumour Hsp90, will they also lose their tumour-cell-specific toxicity? The

study by Kamal *et al.*² suggests that subverting the natural rules of attraction that determine 17-AAG binding to Hsp90 might prove to be counter-productive. Instead of developing drugs with a higher affinity for Hsp90, what might be needed are compounds that have bigger differences in the affinities with which they bind the two forms of Hsp90.

Kamal and colleagues' findings will undoubtedly affect the future design of Hsp90 drugs, but the study also has general implications for anticancer drug development. It suggests that it is not enough to identify a potential molecular target — the drugs directed against that target must also be assessed in an appropriate cellular context. ■

Len Neckers is in the Urologic Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

e-mail: len@helix.nih.gov

Yong-Sok Lee is in the Center for Molecular Modelling, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892, USA.

1. Ferrarini, M., Heltai, S., Zocchi, M. R. & Rugari, C. *Int. J. Cancer* **51**, 613–619 (1992).
2. Kamal, A. *et al. Nature* **425**, 407–410 (2003).
3. Ritossa, F. *Cell Stress Chaperones* **1**, 97–98 (1996).
4. Maloney, A. & Workman, P. *Expert Opin. Biol. Ther.* **2**, 3–24 (2002).
5. Chiosis, G. *et al. Chem. Biol.* **8**, 289–299 (2001).
6. Scheibel, T. & Buchner, J. *Biochem. Pharmacol.* **56**, 675–682 (1998).
7. Banerji, U. *et al. Proc. 93rd Ann. Meet. Am. Assoc. Cancer Res.* **43**, A1352 (2002).
8. Jez, J. M., Chen, J., Rastelli, G., Stroud, R. M. & Santi, D. V. *Chem. Biol.* **10**, 361–368 (2003).
9. Rinehart, K. L. & Shields, L. S. *Fortschr. Chem. Org. Naturstoffe* **33**, 231–307 (1976).
10. Panaretou, B. *et al. Mol. Cell* **10**, 1307–1318 (2002).
11. Riggs, D. L. *et al. EMBO J.* **22**, 1158–1167 (2003).

Applied physics

Spintronics gets a magnetic flute

Jonathan Sun

Magnetic-memory devices of the future could be based on 'spintronics', through switching the directions of electron spins. New work confirms the physics behind a spin-switching mechanism.

Progress in understanding the microscopic behaviour of electrons continues to open up new frontiers for materials and device research, and vice versa. Watching how electrons move, at high current density, through structures that are less than 100 nm in size has revealed a wealth of new physics based on the 'spin' properties of these particles, and

potentially a new class of electronic device: a spin-transfer switch for magnetic-memory 'spintronics'. Experiments reported by Kiselev *et al.*¹, on page 380 of this issue, have now proved the nature of the physics at play.

A magnet usually responds to an electric current because of the magnetic field generated by the current. But if the magnet is

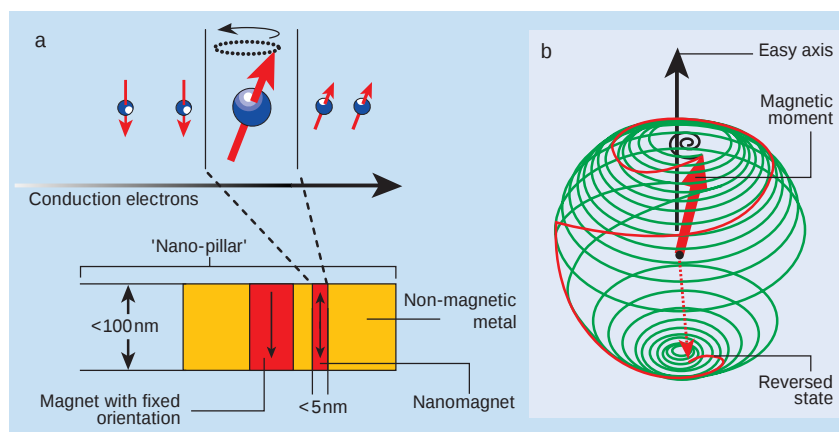


Figure 1 Magnetic excitation and switching through spin transfer. **a**, As conduction electrons pass a magnet, their spins preferentially align in the magnet's direction. As the electrons encounter a nanomagnet, sandwiched between layers of non-magnetic material close to the fixed-orientation magnet, the direction of their spins is repolarized to match that of the nanomagnet. As a result, the nanomagnet's magnetic moment begins to precess, turning like a spinning-top about its axis. **b**, If the current (that is, the rate of electrons passing) is below a threshold value, the nanomagnetic moment relaxes back to its 'easy' axis (black); if the current is just above threshold, the moment follows many cycles of precession until its direction is reversed (green); when the current is well above threshold, the moment quickly reaches its reversed state (red).

smaller than 100 nm or so in size, a new force is expected to emerge^{2–4}. When the electrons that constitute the current pass through a magnetic conductor, their spins will become preferentially aligned to the magnetic direction — that is, they are ‘spin-polarized’. These spins may be repolarized into a new direction when they encounter another magnet (Fig. 1a). In repolarizing the current, the nanomagnet experiences a torque (or turning force) associated with the change in angular momentum that occurs from the rotation of the electron spins. This spin-transfer torque can pump enough energy into the nanomagnet for its magnetic moment to precess — that is, it moves at microwave frequencies around the symmetry axis with ever-increasing amplitude until it reverses its orientation, accomplishing a magnetic switch (Fig. 1b).

The spin-transfer mechanism is both unique and efficient. It is unique because it relies on the amplification of magnetic precession. Under spin-transfer excitation, such precession can even become persistent under balanced conditions with a steady current and a static magnetic field. The mechanism is efficient because, theoretically, it would need a current of only a few hundred microamperes to reverse the moment of a nanomagnet whose reversal would otherwise

require the action of a very strong magnetic field (of the order of a tesla)⁵. This makes the spin-transfer mechanism very interesting as a means of writing magnetic memory.

The current-driven reversal of a nanomagnet has been seen in experiments^{5–8}, with observations that are consistent with a spin-transfer mechanism. The presence of spin-transfer dynamics at microwave frequencies has also been implied by earlier work⁹ in which a spin-transfer junction was irradiated with microwaves. To definitively prove the presence of a spin-transfer mechanism, however, the reverse microwave effect should be sought — that is, the persistent emission of microwaves (generated as the nanomagnet moment precesses) from a device through which current is flowing. This would be a unique signature for spin transfer and impossible for a current-induced magnetic field to produce.

This is exactly what Kiselev *et al.*¹ have done. They attached a broadband microwave receiver to a painstakingly fabricated, magnetic ‘nano-pillar’ — two interleaved layers of magnetic and non-magnetic materials, 70 nm by 130 nm in lateral size. The electron spin direction is set as the current passes through the thicker (fixed) magnetic layer, about 40 nm thick (Fig. 1a). As the current then moves into the much thinner magnetic

layer (3 nm thick), it creates a spin-transfer torque on the magnetic moment of this layer, which emits microwaves as its moment precesses continuously under spin-transfer excitation. By detecting this microwave radiation, Kiselev *et al.* were able to map out the oscillation strength at different frequencies as a function of the electric current and magnetic field, and then to compare this map quantitatively with a model calculation.

It is a very convincing experiment. The only input is a steady electric current. The output undeniably contains microwave emission, in the expected frequency range. Through the spin-transfer mechanism, a current is seen to excite high-frequency magnetic precession, much like the flow of air through a flute excites an audible air vibration.

A magnetic flute sounds like music to the ear for spintronics. What future will it hold for computer memory? In the short term, this will depend critically on two factors. First, the threshold current density needs to be lowered: integration with existing electronic technology would require switching currents of less than 10⁵ to 10⁶ amperes per square centimetre, but for present spin-transfer switches these are more than 10⁷ amperes per square centimetre. Second, read-out signals need to be larger, which means that the resistance of a spin-transfer

Zoology

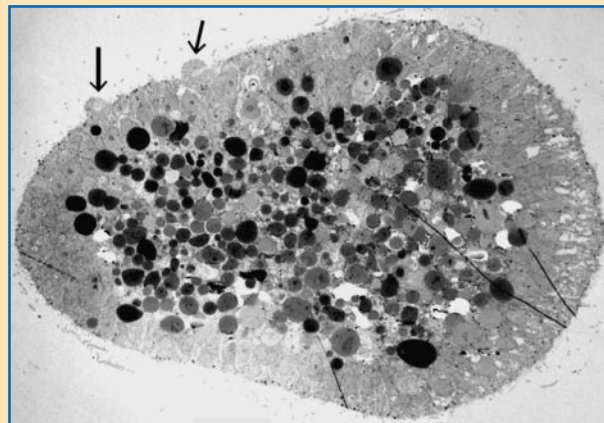
Light touch on the rudder

The larva of the box jellyfish *Tripedalia cystophora* is an uncomplicated creature: it consists of only five cell types. But studies of the animal and of one cell type in particular, the light-sensitive ‘ocellus’, have produced an intriguing set of observations. Most strikingly, the larvae seem not to have a nervous system of any kind to which the photoreceptors could be connected. In the words of Karin Nordström and colleagues, whose work it is, the photoreceptors are “self-contained sensory-motor entities” (*Proc. R. Soc. Lond. B* doi: 10.1098/rspb.2003.2504).

The life cycle of box jellyfish has three phases: a swimming larval stage; a stationary polyp; and the medusa, the familiar jellyfish form, again free-swimming, which in this case is square in cross-section (hence the name of the group). The larva is pear-shaped, but only about 200 μm in length. Tooled up with a transmission electron microscope, Nordström *et al.* set about looking at it in detail.

An individual larva has 10–15 ocelli, two of which are arrowed in this micrograph of the whole organism. Each consists of a single cell, and together they form an array around the rear-end of the animal relative to the direction of movement. The cell contains a ‘cup’ of screening pigment, which is filled with structures that Nordström *et al.* argue are the light-sensing devices. The cups point out at an angle towards the front of the animal. In swimming, the animal also rotates continually (at about two turns per second), so the photoreceptors constitute a scanning system of the light conditions ahead of it.

Another feature of each ocellus is a protruding hair-like cilium. Certain cilia can themselves act as photoreceptors. The authors think that that’s not the case here: such a function requires heavy modification, not evident in the ocellar cilia of the *Tripedalia* larva. So could the cilium be functioning as an active motor? Again no, it seems. Another larval cell type bears cilia for propulsion;



and the ocelli don’t have the extra power packs, in the form of mitochondria, that would be expected in a system that has the dual functions of light-sensing and propulsion.

The explanation that Nordström *et al.* plump for is that the function of the ocellar cilia is that of rudders. They control a larva’s swimming direction by flexing and stretching, steering it towards (or away

from) desirable (or undesirable) conditions signalled by the light regime.

There remains the question of why an array for sensing light conditions ahead of an organism should be set at the back of the body, rather than at the front. Nordström *et al.* have an explanation for that too. It’s because the sensor is also a rudder, and rudders work best at the stern. **Tim Lincoln**

switch must be increased while still keeping its magnetic response at an adequate level. There is no lack of challenges ahead, but Kiselev *et al.*¹ have shown that we are on the right path.

Jonathan Sun is in the IBM Research Division, Thomas J. Watson Research Center, PO Box 218, Yorktown Heights, New York 10598, USA.
e-mail: jonsun@watson.ibm.com

1. Kiselev, S. I. *et al. Nature* **425**, 380–383 (2003).

2. Slonczewski, J. C. *J. Magn. Mag. Mater.* **159**, L1–L7 (1996); **195**, L261–L268 (1999).
3. Berger, L. *Phys. Rev. B* **54**, 9353–9358 (1996).
4. Bazaliy, Y. B., Jones, B. A. & Zhang, S.-C. *Phys. Rev. B* **57**, R3213–R3216 (1998).
5. Sun, J. Z. *J. Magn. Mag. Mater.* **202**, 157–162 (1999).
6. Tsoi, M. *et al. Phys. Rev. Lett.* **80**, 4281–4284 (1998).
7. Wegrowe, J. E., Kelly, D., Jaccard, Y., Guitienne, Ph. & Ansermet, J.-Ph. *Europhys. Lett.* **45**, 626–632 (1999).
8. Katine, J. A., Albert, F. J., Buhrman, R. A., Myers, E. B. & Ralph, D. C. *Phys. Rev. Lett.* **84**, 3149–3152 (2000).
9. Tsoi, M. *et al. Nature* **406**, 46–48 (2000).

Plant biology

Water gate

N. Michele Holbrook and Maciej A. Zwieniecki

Flooding reduces the ability of roots to absorb water. The molecular basis for this paradox involves the regulation of water-channel proteins by the pH inside root cells.

Plants are among the thirstiest of creatures. Of the more than 60 trillion tonnes of water that cycles each year from the land to the atmosphere, nearly two-thirds passes through the bodies of living land plants¹. Plants require these large amounts of water because their mode of acquiring carbon dioxide exposes their interior to the drying power of the atmosphere. Without water, forests become stunted, agricultural yields decline, and house plants shrivel and die. Yet, as is so often the case, too much of a good thing can also be a problem. Flooding typically reduces both photosynthesis and growth, and, in the short term, can even cause plants to wilt. This paradox of famine in the midst of plenty arises from the fact that flooding causes roots to become less permeable to water. On page 393 of this issue², Tournaire-Roux and colleagues elucidate a cascade of molecular events that links the low oxygen levels associated with flooding, via changes in the pH inside cells, to this decrease in root permeability.

The large-scale movement of water from roots to leaves takes place almost entirely through spaces that are not bounded by cell membranes. Only as water enters roots is there substantial evidence that this 'transpiration stream' flows through living cells. The ease with which water can traverse root cells is, to a large degree, governed by the presence of highly selective channel proteins known as aquaporins, specifically those of a subgroup known as 'plasma membrane intrinsic proteins' (PIPs). These proteins are present in the plasma membranes of root cells and enhance the membranes' permeability to water³.

Because of the key role of aquaporins in water uptake by roots⁴, changes in the conductance of individual water channels — a process referred to as 'gating' — are likely to be involved in mediating the effects of flooding. Previous studies have shown that aqua-

porins can be gated by various factors, including enzymatic modification with phosphate groups⁵, intracellular calcium levels^{5,6} and, most recently, protons^{5–8}. But the molecular mechanisms that underlie the regulation of aquaporin permeability — including how it might be connected to flooding — have until now remained elusive.

Gases diffuse around 10,000 times more slowly through water than through air, with the result that roots in flooded soils quickly become exposed to conditions of low oxygen. Early cellular responses to oxygen deprivation include a marked decrease in pH in the cell interior, or cytosol. Tournaire-Roux *et al.*² have now investigated whether these changes are linked to changes in the permeability of roots to water. Using roots of the model plant *Arabidopsis thaliana*, the authors manipulated the pH in the cytosol of root cells in one of three ways: by decreasing the availability of oxygen; by inhibiting respiration; and by loading the roots with acid. In all cases, decreases in cytosolic pH were associated with a decrease in root permeability. Altering the pH outside root cells did not have this effect.

Do decreases in cytosolic pH alter root permeability by affecting the conductance of PIP-type aquaporins? To test this hypothesis, Tournaire-Roux and colleagues inserted aquaporins into the plasma membrane of immature frog eggs. Osmotic swelling assays showed that water influx through most PIP-type aquaporins was selectively blocked by cytosolic acidification. In contrast, the activity of an aquaporin from vacuoles (an intracellular compartment) was insensitive to changes in pH. By using a series of mutant proteins, Tournaire-Roux *et al.* demonstrated that charged amino acids (particularly histidine at position 197) in the cytosolic vestibule of the aqueous pore are responsible for the observed pH sensitivity of the PIPs. These findings provide a detailed

explanation for the mechanisms by which root permeability to water is downregulated in response to anoxia.

What is still unclear, though, is why a plant — flooded or otherwise — would ever choose to restrict water flow through its roots. Roots acclimate to waterlogged soils primarily by forming large gas-filled spaces in the root cortex that provide a low-resistance path for oxygen diffusion⁹. So the physiological significance of Tournaire-Roux and colleagues' findings may lie in the possibility that reductions in root permeability can enhance this morphological acclimation of flooded roots.

This may not be as far-fetched as it seems. Air spaces in the root are formed when specific cells in the root cortex undergo programmed cell death; the cue for this is a build-up of the gaseous hormone ethylene⁹. Flooding leads to an increase in ethylene concentrations in roots, both because anoxia activates an enzyme called ACC synthase, the key biosynthetic enzyme for ethylene, and because waterlogged soils retard the diffusion of the gaseous hormone away from the root. But ethylene accumulation is also affected by the rate at which ACC, the precursor to ethylene, is transported away from the roots in the transpiration stream¹⁰. So downregulation of root permeability could theoretically provide a mechanism that, by reducing the flow of water through roots, accelerates the build-up of ethylene and the transformation of the root into a form that is appropriate for dealing with flooding.

Whether or not this indeed occurs requires further experiments using the sort of genetically modified plants made possible by Tournaire-Roux and colleagues' findings. Plants with pH-insensitive aquaporins could now be engineered, allowing researchers to test whether a decrease in root permeability is required to signal the need for the formation of air spaces. But for now, it is satisfying to see that the paradoxical thirstiness of flooded plants has been placed on firmer ground.

N. Michele Holbrook and Maciej A. Zwieniecki are in the Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA.
e-mail: Holbrook@oeb.harvard.edu

1. Sack, L. & Tyree, M. T. in *Movement of Water and Carbohydrates in Plants: Integration of Long Distance Transport Processes* (eds Holbrook, N. M. & Zwieniecki, M. A.) (Academic, San Diego, in the press).
2. Tournaire-Roux, C. *et al. Nature* **425**, 393–397 (2003).
3. Tyerman, S. D., Niemietz, C. M. & Bramley, H. *Plant Cell Environ.* **25**, 173–194 (2002).
4. Martre, P. *et al. Plant Physiol.* **130**, 2102–2110 (2002).
5. Németh-Cahalan, K. L. & Hall, J. E. *J. Biol. Chem.* **275**, 6777–6782 (2000).
6. Gerbeau, P. *et al. Plant J.* **30**, 71–81 (2002).
7. Yasui, M. *et al. Nature* **402**, 184–187 (1999).
8. Zeuthen, T. & Klaerke, D. A. *J. Biol. Chem.* **274**, 21631–21636 (1999).
9. Buchanan, B. B., Gruissem, W. & Jones, R. L. (eds) *Biochemistry and Molecular Biology of Plants* (Am. Soc. Plant Physiol., Rockville, 2000).
10. Bradford, K. J. & Yang, S. F. *Plant Physiol.* **65**, 322–326 (1980).

Obituary

Edward Teller (1908–2003)



For a man who believed that the huge number of his enemies and critics was a proof of his public influence, Edward Teller, who died of a stroke on 9 September at Palo Alto, California, was oddly thin-skinned: unintended slights could cut him to the quick. Even in his eighties (he was 95 when he died) he craved affection, or at least admiration. Perhaps that was a consequence of his maimed foot, acquired as a graduate student at Munich in an unplanned encounter with a street-tram. But his limp did not prevent him from playing tennis regularly.

Teller was born into a family of assimilated Jews in Budapest on 15 January 1908. His father, the lawyer Max, was sufficiently well-to-do to send his precocious son to private schools — notably to the Minta Gymnasium, one of Budapest's smart secondary schools at the start of the past century.

Three factors shaped the young Teller's fate. For a youth with a mathematical bent, the two great revolutions of the twentieth century in the physical sciences beckoned: he plumped for quantum mechanics in preference to relativity. Then anti-Semitism was on the rise in the Austro-Hungarian empire, meaning that young Jews might be better off elsewhere. The family (Max, in reality) decided that young Edward should become a chemical engineer and packed him off to the Karlsruhe Technische Hochschule, in the Rhine basin. For intending chemical engineers, there was no better place.

Iconoclast from the start, Teller turned away from the job security his father had planned for him and immersed himself in theoretical physics. His undergraduate

Brilliant physicist, who courted controversy and had a big say in US arms policy

degree completed, he embarked on a tour of the intellectual capitals of his chosen field on the European mainland: to Munich (where Arnold Sommerfeld had extended Bohr's — semi-classical — theory of atomic electron orbits to allow for special relativity, and so had made some sense of the spectrum of helium); to Leipzig, where there were Friedrich Hund and the young Werner Heisenberg; and to Göttingen (with Max Born as guru).

Teller's first important paper, published in 1930, was the first calculation, in the then-new language of wave mechanics, of the stationary states of the singly ionized hydrogen molecule. It remains the most vivid illustration of how a single electron can hold two atomic nuclei together. Teller was then Heisenberg's graduate student. The paper ensured him his PhD that year. Three years later, he and Heisenberg published a paper on the likely symmetry-breaking among the nuclei of polyatomic molecules. This idea was carried further in 1937, after Teller's move to the United States, in a paper with Emil Jahn, which eventually led to the explanation of an anomalous band in the ultraviolet spectrum of benzene. The underlying idea of the Jahn–Teller effect is to allow for a distortion of the molecule, whereupon the highly degenerate electronic ground state is not that with the least energy.

Teller left Germany in 1933 when, with

the Nazi rise to power, it had become clear that this, too, was no place to be Jewish. He was rescued by an imaginative scheme cooked up by the Rockefeller Foundation of New York and Niels Bohr at Copenhagen. Bohr had persuaded the foundation that European scientists needed emergency help, that Copenhagen would be a safe staging post and that the foundation should not enforce its rule that fellowships were awarded only to those with somewhere to go permanently, but instead to leave that to him (Bohr).

At Copenhagen, Teller met George Gamow, then a professor at George Washington University, who offered him a permanent job. Meanwhile Bohr, faithful to his pact with Rockefeller, had found him a place at University College London. Teller took that post for just a year (1934–35), then moved to the United States and promptly applied for naturalization papers. At George Washington University, he began a collaboration that led to a refinement of Enrico Fermi's theory of β -decay. The essential idea was to allow for the interaction between the spins of the neutrinos emitted from decaying nuclei and the spins of the heavy particles in the disintegrating nucleus, leading to selection rules differing from those of the simple Fermi theory.

Then came the outbreak of the Second World War, and the rest, as they say, is history. But what history! With the Manhattan Project — the programme to build the atom bomb — under way, Teller seems to have been recruited by his fellow-Hungarian Leo Szilard, to join Fermi, first at Columbia University and then at the University of Chicago. By 1944, they were both installed at the Los Alamos facility in New Mexico, established in 1943. Teller was one of a handful of people in the theoretical department, with Hans Bethe at its head, but with the likes of Rudolf Peierls as occasional colleagues.

The resulting achievement, the atom bomb, was first exploded in a test at Alamogordo and then to horrendous effect over Hiroshima and Nagasaki.

Trouble, and the roots of Teller's estrangement from much of the scientific community, flared up only in April 1954, when the Personnel Security Board (PSB) of the US Atomic Energy Commission started closed hearings on the question of whether Robert Oppenheimer, director of Los Alamos since it began until 1945, and then director of the Institute of Advanced Study in Princeton, should be stripped of the security clearance that allowed him access to secret information.

Since his amicable return to academic life, Oppenheimer had been employed as a

part-time consultant by the Atomic Energy Commission and headed its General Advisory Committee. But with Senator Joe McCarthy and his House Un-American Activities Committee carrying a public pillory around the United States, J. Edgar Hoover at the Federal Bureau of Investigation had taken a look at Oppenheimer's file and taken a dim view of what he saw. Matters came to a head in 1953, when Oppenheimer was interviewed by the chairman of the Atomic Energy Commission, Lewis Strauss, and given a day to decide whether to resign or face the indignity of a full-scale hearing.

The charge-sheet (but the PSB insisted that there were no "charges", only "matters for concern") consisted mostly of reports of contacts with members of the Communist Party, and the failure to report, or delay in reporting, attempts to solicit secret information. But hidden among its tittle-tattle is the charge that Oppenheimer had sought to delay the development of the hydrogen bomb — the 'super'.

The public record of the PSB's proceedings are an evocation not only of the early 1950s in the United States and the fevered paranoia engendered by McCarthy, but also of the peculiar fondness in the United States (and elsewhere now) for using bureaucratic quasi-legal procedures to determine shades of opinion.

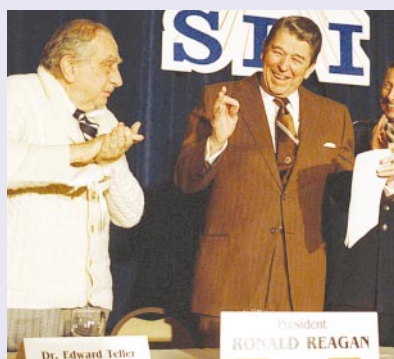
Teller was forthright at the hearing. He had known about the feasibility of fission bombs since 1939 (as had many technical people). Teller added that he also knew that hydrogen bombs were potentially a more powerful source of explosive energy than fission bombs, and that, from his time at Chicago until Hiroshima, he and his colleagues had talked about their feasibility. From the beginning of Los Alamos, the hydrogen bomb was partly a continuing seminar conducted by Teller's group, partly a fall-back weapon if fission bombs should come to nothing — despite the fact that a fission bomb was required to ignite a 'super'.

And what of Oppenheimer as a security risk? Teller's opinion was this: "In a great number of cases, I have seen Dr Oppenheimer act — I understand that Dr Oppenheimer acted — in a way which was for me exceedingly hard to understand. I thoroughly disagreed with him in numerous issues and his actions frankly appeared to me to be confused and complicated. To this extent I would like to see the vital interests of this country in hands which I understand better and therefore trust more."

To be fair to Teller, he was not the only scientist who shafted Oppenheimer. Another was the late Luis Alvarez, who disagreed with an opinion said to have been expressed by Oppenheimer that if the United States

did not build thermonuclear weapons, the Russians would not either. The difficulty with quasi-judicial proceedings is that they thrive on hearsay evidence.

Bethe's evidence at the hearing was revealing of Teller's uncollegial turn of mind. When fission weapons were everybody's goal, and when the results of experiments or calculations had to be produced in predetermined sequence, Teller would not keep to the timetable. The end result was that Teller and his small



group were cast adrift, allowed to pursue their own interests while weakening the common enterprise.

In his later life, Teller would shrug off the epithet 'Father of the H-bomb', but he was busily earning it from 1942 until a decade later, when the first US thermonuclear bomb was exploded at Eniwetok atoll in the Pacific.

After the Second World War, Teller briefly considered running as US senator for California (in the Republican cause), when a waggish colleague told him that he enjoyed more power as he was than any politician. But Teller was already a politician: he exercised great charm when it suited him, lobbied hard for his favourite causes and knew how to manipulate people to act against the grain.

That is why he never understood why his espousal of the H-bomb, and his damning of Oppenheimer with faint praise, caused widespread (if not universal) distaste in the scientific community. Louis Rosen and his wife, who had been neighbours of the Tellers when they lived on the site, heard that Teller was paying a brief visit to Los Alamos and invited all the people they could think of to a kind of 'Welcome back!' party. Mrs Rosen told me a few years ago that nobody (except Teller) showed up. "He hung about for half an hour and then just walked off," she said.

My own encounters with him were at first stormy. By 1960, Teller was doing propaganda for his belief that infantrymen should have nuclear weapons in their knapsacks so as to practise individual deterrence. I was then working at the *The*

Washington Post, on exchange from the *Manchester Guardian* (as it was), when Teller came to lunch.

As the lunch broke up, Teller said to me, "Can I have a private word with you?" When the others had left, he stood directly in front of me, as if ready for a fist fight, and demanded, "What colour are my eyes?" "Grey," I said, but I knew instantly what he meant. In a piece I had written, probably two years earlier, I remembered that I had described him as "black-eyed Teller". He explained that he did not care on his own account: it was his wife who had been hurt.

The next meeting was even briefer. Teller was in London promoting a book, his agent had offered him for an interview and I had been asked to do the interviewing. The agent put his head through the studio door first, saying that Teller would be just a minute, when indeed he also appeared. Our eyes met briefly, he shouted, "It's you!", turned on his heel and was not seen again.

Surprisingly, then there came the charm, first in a call from California in about 1984. The voice was as gravely as Henry Kissinger's, but more heavily accented. Could I have breakfast with him on a particular day the following week? I duly turned up at his hotel. He was then promoting *Star Wars* (Reagan edition), and wanted to talk me out of the hostile view expressed in a leading article in *Nature*. We agreed to disagree, but he suggested we meet again when he was next in London.

That time was another breakfast, but as I should have guessed, he now had a manuscript for publication, likely to fill 20 pages of *Nature*. It was all about 'brilliant pebbles', the 50-kg laser satellites launched in swarms and each capable of knocking out a hostile missile without further intervention. I explained about referees and he offered a shorter version. We then had a civilized conversation. I never heard from him again. The revised version was published, but in the expectation (mine) that there would be a flood of critical comment. There was hardly any.

What Teller's friends say of him is true: his mind was quick and ingenious. His imagination was breathtaking. His physical energy was prodigious. His intellectual courage was formidable — he invited brickbats. His detractors, probably the majority, are also correct in saying that he was devious and manipulative, that his intellectual courage verged on the foolhardy or provocative, that his intolerance for those who failed to agree with him was inexcusable. All agree that he was an exceptionally gifted physicist. John Maddox
John Maddox, now Emeritus Editor, was Editor of *Nature* for the periods 1966–73 and 1980–95.

Developmental biology

Protein destruction from egg to embryo

Dev. Cell **5**, 451–462 (2003)

As a fertilized egg develops into an embryo, protein levels change dramatically. Yet gene transcription — the initial step in protein synthesis that makes an RNA copy of a DNA-based gene — is not needed. So how does the embryo manage to develop correctly?

Jason Pellettieri *et al.* offer a suggestion. They show that, in the nematode worm *Caenorhabditis elegans*, the MBK-2 protein coordinates the breakdown of several maternal proteins. When the process is impaired, fertilized eggs fail to undergo division by mitosis (which ordinarily generates cells that have two copies of each chromosome). The eggs also lack polarity — a sense of orientation — and so fail to lay down the beginnings of a head-to-tail body axis.

The distribution of MBK-2 changes radically after fertilization, so the authors conclude that the protein may control several events around this time. They suggest that it may help to transform a symmetrical, single-celled egg into an asymmetric, patterned embryo. **Helen R. Pilcher**

Environmental science

Lead, leaching and landfill

Environ. Sci. Tech. doi:10.1021/es034155t (2003)

The mountains of electronic waste, or e-waste — televisions, computers, cell phones, and so forth — that are accumulating in industrialized societies pose a potential hazard. They contain toxic heavy metals, such as lead in batteries and solder, which may be leached into groundwater when e-waste is dumped in landfills. But Yong-Chul Jang and Timothy G. Townsend claim that the release of heavy metals may not be as great as suggested by the standard US assessment protocol.

In the United States, lead leaching from

solid wastes is determined using the Environmental Protection Agency's Toxicity Characteristic Leaching Procedure (TCLP). Tested in this way, lead leaching from cathode-ray tubes in televisions and computer monitors exceeds the specified limit for non-hazardous waste. Nevertheless, cathode-ray tubes can still be dumped in municipal landfills in some US states. Jang and Townsend show that when cathode-ray tubes and printed circuit boards in Florida landfill sites were leached using the actual runoff fluid from these sites, rather than the leachate (acetic acid) specified in the TCLP, the amount of lead released was about two orders of magnitude lower. For lead at least, TCLP therefore seems to be not merely a conservative test but potentially a misleading one.

Philip Ball

Biophysics

Cell mail delivered by light

Appl. Phys. Lett. **83**, 2468–2470 (2003)

At present, the usual way to transfer material into a specific living cell is to inject it through a micropipette. This is how genetic material is transmitted during cloning. Kenich Yoshikawa and colleagues demonstrate an alternative method. They use optical tweezers — laser beams that create an optical trap — to pull micrometre-sized particles through the cell walls of plant (ginkgo and cabbage) cells. This is less disruptive than puncturing the membranes with a micropipette.

Optical tweezers can be used to pick up individual DNA molecules if the strands are compacted by complexation agents such as certain polymers. But the grip is too loose for reliable transfer. So the researchers encapsulated the DNA in the nanoscale pores of aluminosilicate zeolite particles, which can be trapped about 40 times more efficiently. They also inserted a calcium-sensitive fluorescent dye into cells using a zeolite vehicle. Ideally, they say, one should find a vehicle that dissolves after delivery.

Philip Ball

Astrophysics

Dark interactions

Preprint astro-ph/0309303

Invisible dark matter, of unknown origin, dominates the mass of galaxies and clusters of galaxies. It cannot be imaged directly. But M. Markevitch and colleagues have measured how strongly dark-matter particles interact with one another — an issue that greatly exercises astrophysicists — by combining data on a one-off event.

The gravity shadow of a dark-matter bullet, speeding through galaxy cluster 1E0657-56, was spotted by other observers in the distortion pattern of faint galaxies lying beyond cluster 1E0657-56. Earlier X-ray observations by Markevitch *et al.* had revealed that hot gas trails in the bullet's wake. The blob of dark matter is fast-moving and sweeps galaxies along with it, but the hot gas travels more slowly. The disparity in velocity can be explained if the probability of dark-matter particles colliding with one another is extremely low, much lower than for gas particles.

Markevitch *et al.* calculate that two 1-gram particles of dark matter must come within an area of 1 cm² of each other to interact. This would rule out stronger interactions, which have been invoked to explain why dark matter in galaxy cores is less concentrated than computer simulations predict.

Joanne Baker

Ecology

Oxygen needs in the corals

Proc. R. Soc. Lond. B (Suppl.)
doi:10.1098/rsbl.2003.0087 (2003)

Coral reef fish can tolerate low levels of oxygen, Göran E. Nilsson and Sara Östlund-Nilsson have discovered. Reef fish were not previously known for their ability to survive in such hypoxic conditions — nor are reefs usually thought of as being low in oxygen.

The researchers looked at 31 species of fish living in the corals of a lagoon on the Great Barrier Reef. All of the fish showed a critical oxygen concentration of around 30% or lower of air saturation. In other words, the fish can maintain a high rate of oxygen consumption, and respire aerobically, until oxygen concentrations drop to these severely hypoxic levels.

Nilsson and Östlund-Nilsson suggest that hypoxia tolerance might help fish to live alongside oxygen-consuming coral. In a lagoon cut off from the sea at low tide — or deep within the reef, where many fish seem to spend the night — respiring coral might create a nocturnal oxygen shortage. The authors also propose that, given the high biodiversity of coral reefs, there may be many unusual adaptations to hypoxia waiting to be discovered.

John Whitfield



B. BISSON/CORBIS SYGMA

Anthropogenic carbon and ocean pH

The coming centuries may see more ocean acidification than the past 300 million years.

Most carbon dioxide released into the atmosphere as a result of the burning of fossil fuels will eventually be absorbed by the ocean¹, with potentially adverse consequences for marine biota^{2–4}. Here we quantify the changes in ocean pH that may result from this continued release of CO₂ and compare these with pH changes estimated from geological and historical records. We find that oceanic absorption of CO₂ from fossil fuels may result in larger pH changes over the next several centuries than any inferred from the geological record of the past 300 million years, with the possible exception of those resulting from rare, extreme events such as bolide impacts or catastrophic methane hydrate degassing.

When carbon dioxide dissolves in the ocean it lowers the pH, making the ocean more acidic. Owing to a paucity of relevant observations, we have a limited understanding of the effects of pH reduction on marine biota. Coral reefs², calcareous plankton³ and other organisms whose skeletons or shells contain calcium carbonate may be particularly affected. Most biota reside near the surface, where the greatest pH change would be expected to occur, but deep-ocean biota may be more sensitive to pH changes⁴.

To investigate the effects of CO₂ emissions on ocean pH, we forced the Lawrence Livermore National Laboratory ocean general-circulation model⁵ (Fig. 1a) with the pressure of atmospheric CO₂ (pCO₂) observed from 1975 to 2000, and with CO₂ emissions from the Intergovernmental Panel on Climate Change's IS92a scenario¹ for 2000–2100. Beyond 2100, emissions follow a logistic function for the burning of the remaining fossil-fuel resources (assuming 5,270 gigatonnes of carbon (GtC) in 1750; refs 6, 7). Simulated atmospheric CO₂ exceeds 1,900 parts per million (p.p.m.) at around the year 2300. The maximum pH reduction at the ocean surface is 0.77; we estimate, using a geochemical model^{8,9}, that changes in temperature, weathering and sedimentation would reduce this maximum reduction by less than 10%.

A review¹⁰ of estimates of palaeo-atmospheric CO₂ levels from geochemical models, palaeosols, algae and forams, plant stomata and boron isotopes concluded that there is no evidence that concentrations were ever more than 7,500 p.p.m. or less than 100 p.p.m. during the past 300 million years (Myr). Moreover, the highest concentrations inferred from the geological record were thought to have developed over many millions of years owing to slow processes involving tectonics and biological evolution.

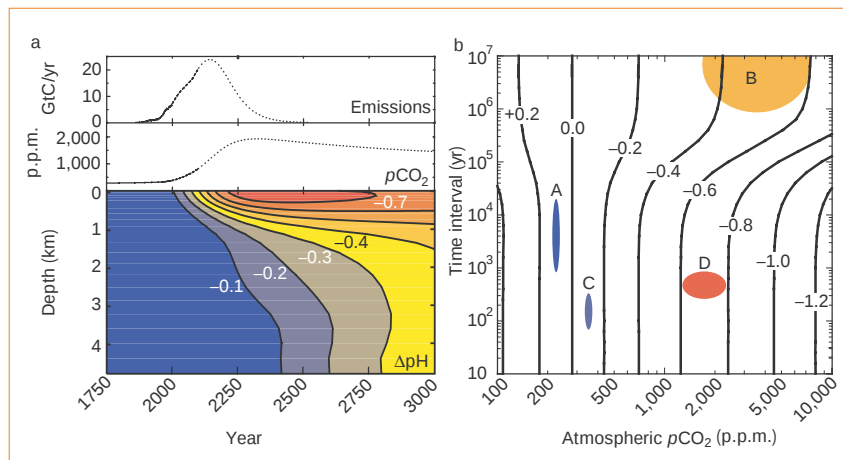


Figure 1 Atmospheric release of CO₂ from the burning of fossil fuels may give rise to a marked increase in ocean acidity. **a**, Atmospheric CO₂ emissions, historical atmospheric CO₂ levels and predicted CO₂ concentrations from this emissions scenario, together with changes in ocean pH based on horizontally averaged chemistry. **b**, Estimated maximum change in surface ocean pH as a function of final atmospheric CO₂ pressure, and the transition time over which this CO₂ pressure is linearly approached from 280 p.p.m. A, glacial–interglacial CO₂ changes¹³; B, slow changes over the past 300 Myr; C, historical changes¹ in ocean surface waters; D, unabated fossil-fuel burning over the next few centuries.

We estimated the effect of past changes in atmospheric CO₂ levels on ocean pH by using a four-box ocean/atmosphere model^{8,9}. Modelled processes include weathering of carbonate and silicate minerals on land, production of shallow-water carbonate minerals, production and oxidation of biogenic organic carbon, production and dissolution of biogenic carbonate minerals in the ocean, air–sea gas exchange of carbon, and transport by advection, mixing and biological processes.

In a series of simulations, atmospheric pCO₂ was varied linearly from the pre-industrial value (about 280 p.p.m.) to stabilization values from 100–10,000 p.p.m. over time intervals of 10–10⁷ yr. For each simulation, we recorded the maximum predicted perturbation in pH in the surface-ocean boxes (Fig. 1b). When a CO₂ change occurs over a short time interval (that is, less than about 10⁴ yr), ocean pH is relatively sensitive to added CO₂. However, when a CO₂ change occurs over a long time interval (longer than about 10⁵ yr), ocean chemistry is buffered by interactions with carbonate minerals, thereby reducing sensitivity to pH changes¹¹.

Based on the record¹⁰ of atmospheric CO₂ levels over the past 300 Myr and our geochemical model^{8,9}, there is no evidence that ocean pH was more than 0.6 units lower than today. Our general circulation model results indicate that continued release of fossil-fuel CO₂ into the atmosphere could lead to a pH reduction of 0.7 units. We

conclude that unabated CO₂ emissions over the coming centuries may produce changes in ocean pH that are greater than any experienced in the past 300 Myr, with the possible exception of those resulting from rare, catastrophic events in Earth's history^{8,12}.

Ken Caldeira*, Michael E. Wickett†

**Energy and Environment Directorate and
†Center for Applied Scientific Computing, Lawrence
Livermore National Laboratory, 7000 East Avenue,
Livermore, California 94550, USA
e-mail: kenc@llnl.gov*

- Houghton, J. T. et al. (eds) *Climate Change 2001: The Scientific Basis* (Contribution of Working Group I to the Third Assessment Report of the IPCC, Cambridge Univ. Press, Cambridge, 2001).
- Kleypas, J. A. et al. *Science* **284**, 118–120 (1999).
- Riebesell, U. et al. *Nature* **407**, 364–367 (2000).
- Seibel, B. A. & Walsh, P. J. *Science* **294**, 319–320 (2001).
- Caldeira, K. & Duffy, P. B. *Science* **287**, 620–622 (2000).
- Metz, B., Davidson, O., Swart, R. & Pan, J. (eds) *Climate Change 2001: Mitigation* (Contribution of Working Group III to the Third Assessment Report of the IPCC, Cambridge Univ. Press, Cambridge, 2001).
- Marland, G., Boden, T. A. & Andres, R. J. Global, regional, and national CO₂ emissions. *Trends: A Compendium of Data on Global Change* http://cdiac.esd.ornl.gov/trends/emis/tre_glob.htm (2002).
- Caldeira, K. & Rampino, M. R. *Paleoceanography* **8**, 515–525 (1993).
- Caldeira, K. & Rau, G. H. *Geophys. Res. Lett.* **27**, 225–228 (2000).
- Crowley, T. J. & Berner, R. A. *Science* **292**, 780–781 (2001).
- Caldeira, K., Berner, R., Sundquist, E. T., Pearson, P. N. & Palmer, M. R. *Science* **286**, 2043 (1999).
- Beerling, D. J. & Berner, R. A. *Global Biogeochem. Cycles* **16**, 101–113 (2002).
- Sanyal, A., Hemming, N. G., Hanson, G. N. & Broecker, W. S. *Nature* **373**, 234–236 (1995).

Competing financial interests: declared none.

Archaeology

Sharp shift in diet at onset of Neolithic

The introduction of domesticated plants and animals into Britain during the Neolithic cultural period between 5,200 and 4,500 years ago is viewed either as a rapid event¹ or as a gradual process that lasted for more than a millennium². Here we measure stable carbon isotopes present in bone to investigate the dietary habits of Britons over the Neolithic period and the preceding 3,800 years (the Mesolithic period). We find that there was a rapid and complete change from a marine- to a terrestrial-based diet among both coastal and inland dwellers at the onset of the Neolithic period, which coincided with the first appearance of domesticates. As well as arguing against a slow, gradual adoption of agriculture and animal husbandry by Mesolithic societies, our results indicate that the attraction of the new farming lifestyle must have been strong enough to persuade even coastal dwellers to abandon their successful fishing practices.

Stable carbon isotopes in human bone collagen act as indicators of past dietary intake³ because marine and terrestrial dietary proteins leave different 'signatures'⁴. Consumption of cereal crops that use the C₃ photosynthetic pathway and of farmed animals should result in a 'terrestrial' bone-collagen carbon-isotope signature ($\delta^{13}\text{C} = -20 \pm 1\text{‰}$, where $\delta^{13}\text{C}$ represents the $^{13}\text{C}/^{12}\text{C}$ ratio), whereas marine foods give a much higher ^{13}C content ($\delta^{13}\text{C} = -12 \pm 1\text{‰}$).

Archaeological evidence for the use of marine foods during the British Mesolithic is limited because very few coastal sites survived the rising sea levels of the more recent Holocene epoch. Some of the best-known exceptions are the late Mesolithic shell middens of western Scotland^{5,6}. Other areas of Atlantic Europe, most notably southern Scandinavia and Brittany, present strong archaeological and isotopic evidence of marine-based economies at this time.

We have measured and collated⁶⁻⁸ the carbon-isotope values of bone collagen of 164 early Neolithic (5,200–4,500 yr BP) and 19 Mesolithic (9,000–5,200 yr BP) British humans. The Neolithic sample is derived from a range of contexts, including causewayed enclosures, chambered tombs, caves and stray finds, from both inland and coastal locations. Although individuals from inland and putative 'élite' contexts are more prominently represented, the results from all of these contexts are unanimous.

Figure 1 shows that, with few exceptions, individuals living near the coast in the Mesolithic show a moderate-to-strong marine isotope signal (for four humans from two inland sites, $\delta^{13}\text{C} = -19.6 \pm 0.8$; for fifteen humans from eight coastal sites, $\delta^{13}\text{C} = -16.2 \pm 2.8$), and that all of the Neolithic humans show a strongly terrestrial isotope signal (for 99 humans from 25 inland sites, $\delta^{13}\text{C} = -20.7 \pm 0.7$; for 68 humans from 19 coastal sites, $\delta^{13}\text{C} = -20.8 \pm 0.7$). These data are comparable with results obtained in Denmark, which also show a rapid dietary change in humans between the Mesolithic and Neolithic at about the same time^{9,10}.

From our findings, we conclude that there was a sudden and marked dietary shift associated with the onset of the Neolithic period in Britain, arguing against a gradual uptake of domesticated plants and animals into Mesolithic society². Marine foods, for whatever reason, seem to have been comprehensively abandoned from the beginning of the Neolithic in Britain.

Michael P. Richards*, Rick J. Schulting†, Robert E. M. Hedges‡

*Department of Archaeological Sciences, University of Bradford, Bradford BD7 1DP, UK
e-mail: m.p.richards@bradford.ac.uk

†School of Archaeology and Palaeoecology, Queen's University Belfast, Belfast BT7 1NN, UK

‡Research Laboratory for Archaeology and the History of Art, University of Oxford, Oxford OX1 3QJ, UK

1. Childe, V. G. *Man Makes Himself* (Watts, London, 1936).
2. Dennell, R. W. *European Economic Prehistory* (Academic, London, 1983).
3. Schwarcz, H. & Schoeninger, M. *Yb. Phys. Anthropol.* **34**, 283–321 (1991).
4. Schoeninger, M., DeNiro, M. & Tauber, H. *Science* **220**, 1381–1383 (1983).
5. Mellars, P. A. *Excavations on Oronsay* (Edinburgh Univ. Press, Edinburgh, 1987).
6. Schulting, R. J. & Richards, M. P. *Eur. J. Archaeol.* **5**, 147–189 (2002).
7. Richards, M. P. & Hedges, R. E. M. *Antiquity* **73**, 891–897 (1999).
8. Schulting, R. J. & Richards, M. P. *Antiquity* **76**, 1011–1025 (2002).
9. Tauber, H. *Nature* **292**, 332–333 (1981).
10. Richards, M. P., Price, T. D. & Koch, E. *Curr. Anthropol.* **44**, 288–294 (2003).

Competing financial interests: declared none.

corrigendum

Insecticide resistance in mosquito vectors

M. Weill, G. Lutfalla, K. Mogensen, F. Chandre, A. Berthomieu, C. Berticat, N. Pasteur, A. Phillips, P. Fort, M. Raymond

Nature **423**, 136–137 (2003)

It has been drawn to our attention (I. Gould and P. A. Zimmerman) that the numbering of the amino-acid sequences and exon presented in our communication differs from that of the corresponding EMBL/GenBank entries. Because mosquito proteins differ in length, the first published acetylcholinesterase three-dimensional structure (from *Torpedo californica*) was used to number structurally identical residues. The glycine residue whose substitution by serine confers insecticide insensitivity was therefore numbered 119, whereas it corresponds to amino acids 247 and 280 of *Culex pipiens* (entry CAD33707) and *Anopheles gambiae* (entry CAD56157) proteins, respectively. In *Anopheles*, this position lies within the third coding exon (exon 5). Our conclusions are not affected by this inconsistency.

brief communications is intended to provide a forum for brief, topical reports of general scientific interest and for technical discussion of recently published material of particular interest to non-specialist readers (communications arising). Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on Nature's website (www.nature.com/nature).

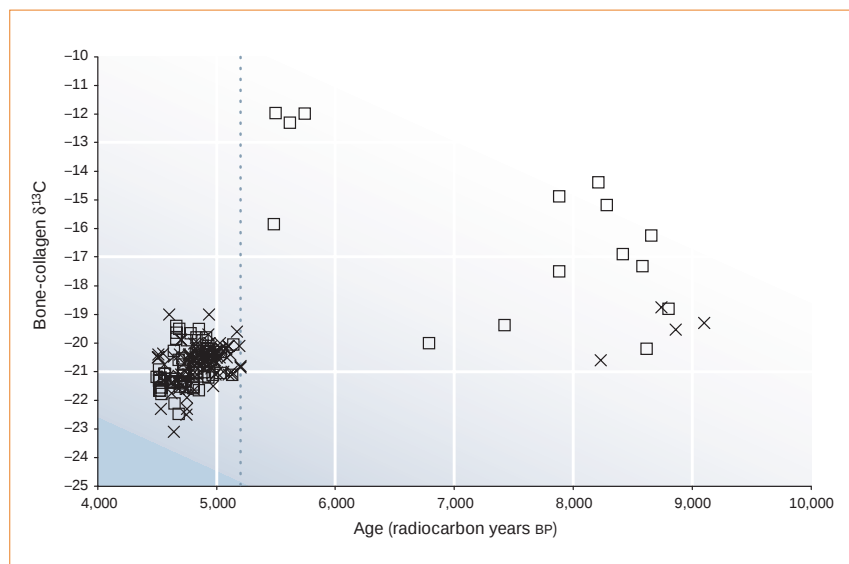


Figure 1 Bone-collagen carbon-isotope ratios and radiocarbon ages of 183 Mesolithic and Neolithic humans from coastal (that is, living within 10 km of contemporary coastline; squares) and inland sites (crosses) in Britain. There is a sharp change in the carbon-isotope ratio at around 5,200 yr BP (about 4,000 calendar yr ac; dotted line) from a diet consisting of marine foods to one dominated by terrestrial protein. This period coincides with the onset of the Neolithic period in Britain. Because of uncertainties in the size of the marine reservoir effect on radiocarbon dates, we have not attempted to calibrate the data here; however, for individuals with typical marine $\delta^{13}\text{C}$ values (about -12‰), radiocarbon dates should be corrected by roughly 400 radiocarbon years.

Bending-related faulting and mantle serpentinitization at the Middle America trench

C. R. Ranero¹, J. Phipps Morgan¹, K. McIntosh² & C. Reichert³

¹GEOMAR and SFB574, Wischhofstrasse 1-3, 24148 Kiel, Germany

²Institute of Geophysics, University of Texas at Austin, 4412 Spicewood Springs Rd., Bldg 600, Austin, Texas 78759-8500, USA

³BGR, Bundesanstalt für Geowissenschaften und Rohstoffe, Stilleweg 2, 30655 Hannover, Germany

The dehydration of subducting oceanic crust and upper mantle has been inferred both to promote the partial melting leading to arc magmatism and to induce intraslab intermediate-depth earthquakes, at depths of 50–300 km. Yet there is still no consensus about how slab hydration occurs or where and how much chemically bound water is stored within the crust and mantle of the incoming plate. Here we document that bending-related faulting of the incoming plate at the Middle America trench creates a pervasive tectonic fabric that cuts across the crust, penetrating deep into the mantle. Faulting is active across the entire ocean trench slope, promoting hydration of the cold crust and upper mantle surrounding these deep active faults. The along-strike length and depth of penetration of these faults are also similar to the dimensions of the rupture area of intermediate-depth earthquakes.

Since the proposition that intermediate-depth earthquakes occur by reactivation of outer rise faults¹, considerable discussion has concentrated on possible mechanisms that would allow seismogenic slip at these large confining pressures where plastic deformation would be anticipated². A commonly invoked mechanism to promote brittle failure is the decrease in effective normal stress induced by increased pore pressure during dehydration reactions^{3–6}. As oceanic plates dive into subduction zones, progressive metamorphism releases fluids chemically bound in the slab. Such fluid release, if concentrated along pre-existing fault weaknesses, may facilitate seismogenic slip across the upper part of the slab⁴ or even tens of kilometres into the slab's mantle⁷. In Tonga, the similar orientation and dip (relative to the slab surface) of nodal planes of intraslab intermediate-depth seismicity and nodal planes of outer rise events has been interpreted to support this hypothesis⁸. Although bending-related normal faulting occurs at all subduction zones, little high-resolution data exists to permit testing of this hypothesis.

If slab-released fluids enter overlying asthenosphere, there they can induce 'wet' partial melting and the formation of magmatic arcs⁹. It has been realized that at some subduction zones, dehydration of most of the hydrous phases contained in subducting sediments and ocean crust should occur at shallower depths than the depths of magma generation—thus the fluids inducing arc-melting are likely to come from the slab's mantle¹⁰. However, the feasibility of large-scale hydration of the upper ocean mantle is still under debate¹¹, with different calculations of the global fluid budget entering subduction zones, assuming that water might be stored in the sediments and crust^{12,13} or in the mantle⁶. Where most of the hydration of the incoming plate takes place also remains in doubt: near a spreading centre^{4,10}, intraplate by fluids rising from underlying mantle plumes¹⁴, or beneath the outer rise, mainly in the crust⁴, or perhaps also in the mantle⁷. At the outer rise, some earthquakes have been proposed to cut >20 km deep into the lithosphere^{15,16}, suggesting a possible pathway for sea water to reach and react with cold lithospheric mantle. (However, the typical hypocentral depths of outer rise earthquakes remains under debate.)

Here we present new evidence documenting that bending-related faulting at the trench offshore Nicaragua has created a pervasive extensional faulting system that cuts through the crust of this

incoming plate, penetrating deeply into the lithospheric mantle. Faulting remains active across the entire ocean trench slope, promoting deep percolation of water and modification of the properties of the incoming ocean plate just before its subduction.

Relation between bend-faulting and seafloor spreading fabric

Multibeam bathymetry along the Middle America trench (MAT) shows that bending-related extensional faulting (bend-faulting) is pervasive across most of the ocean trench slope (Fig. 1). The crustal thickness and the orientation of the tectonic fabric formed at the spreading centre seem to govern the amount of faulting and trench depth for each of the four tectonic segments within Fig. 1. Along this margin, the age of the subducting lithosphere most probably has a secondary role as it changes little and gradually along the incoming plate—lithosphere formed at the Galapagos spreading centre (segments 1–3) ranges from ~14 Myr off South Costa Rica to ~22 Myr off the middle of the Nicoya peninsula, while lithosphere formed along the East Pacific Rise (segment 4) is 24 Myr along the Nicaragua trench¹⁷. Segments with the most pervasive faulting and larger offsets (segments 2 and 4 in Fig. 1) are associated with 5–7-km-thick crust and magnetic anomalies striking at low angles to the trench.

This indicates that, for lithosphere of the same age, bend-faults develop best when they reactivate faults formed at the palaeospreading centre, which occurs when the strike of the inherited tectonic fabric is nearly parallel to the axis of bending of the incoming plate. Feedback between faulting and flexure is inferred from the structures: faults are initially reactivated owing to bending, but trench parallel faulting appears to lead to larger bending and to the development of new faults, in spite of a similar 'elastic thickness' across much of the margin. This is best seen by comparing the ocean plate facing the Nicoya peninsula and Nicaragua, where the age and crustal thickness are essentially the same, but the number of faults and their offsets are larger to the north, where bending is also larger.

Similarly, the plate at segment 2 bends more than at segment 3, where magnetic anomalies strike perpendicular to the trench. The subducted slab steepens gradually to the north from 35–45° beneath Costa Rica to ~65° beneath Nicaragua¹⁸, implying increasing net plate bending towards the North. We concentrate our analysis on segment 4 (Fig. 1), which shares the morphologic characteristics of

many circumpacific subduction zones in regions without oceanic plateau and spreading-centre subduction (for example, Peru and Chile trench^{19–21}; Japan-Kuril trench²²; the Tonga-Kermadec, Aleutian and Java trenches²³).

Seismic images of bend-faulting

Multichannel seismic (MCS) reflection images across the oceanic trench slope show two main sets of reflections beneath the sediments. A set of subhorizontal reflections occurs ~ 1.7 s beneath the top basement delineating the base of the crust (Fig. 2a, b). A second set appears as trenchward-dipping reflections across the crust and

mantle. Some dipping events traverse the ~ 5.5 -km-thick crust and align with reflections in the mantle which extend at least to depths of 18–20 km beneath the sea floor. Deeper than 18–22 km, reflections are obscured by the energy of the water-layer multiple so that their ultimate depth is unconstrained. Locally, there are also seaward-dipping reflections in the mantle (Fig. 2a), although this reflectivity is notably less pervasive than the landward-dipping fabric. The seismic profiles shown in Fig. 2a and b were shot in opposite directions, normal to the seafloor topography, indicating that neither the reflections nor their dominant dipping direction are artefacts of the data acquisition and processing²⁴.

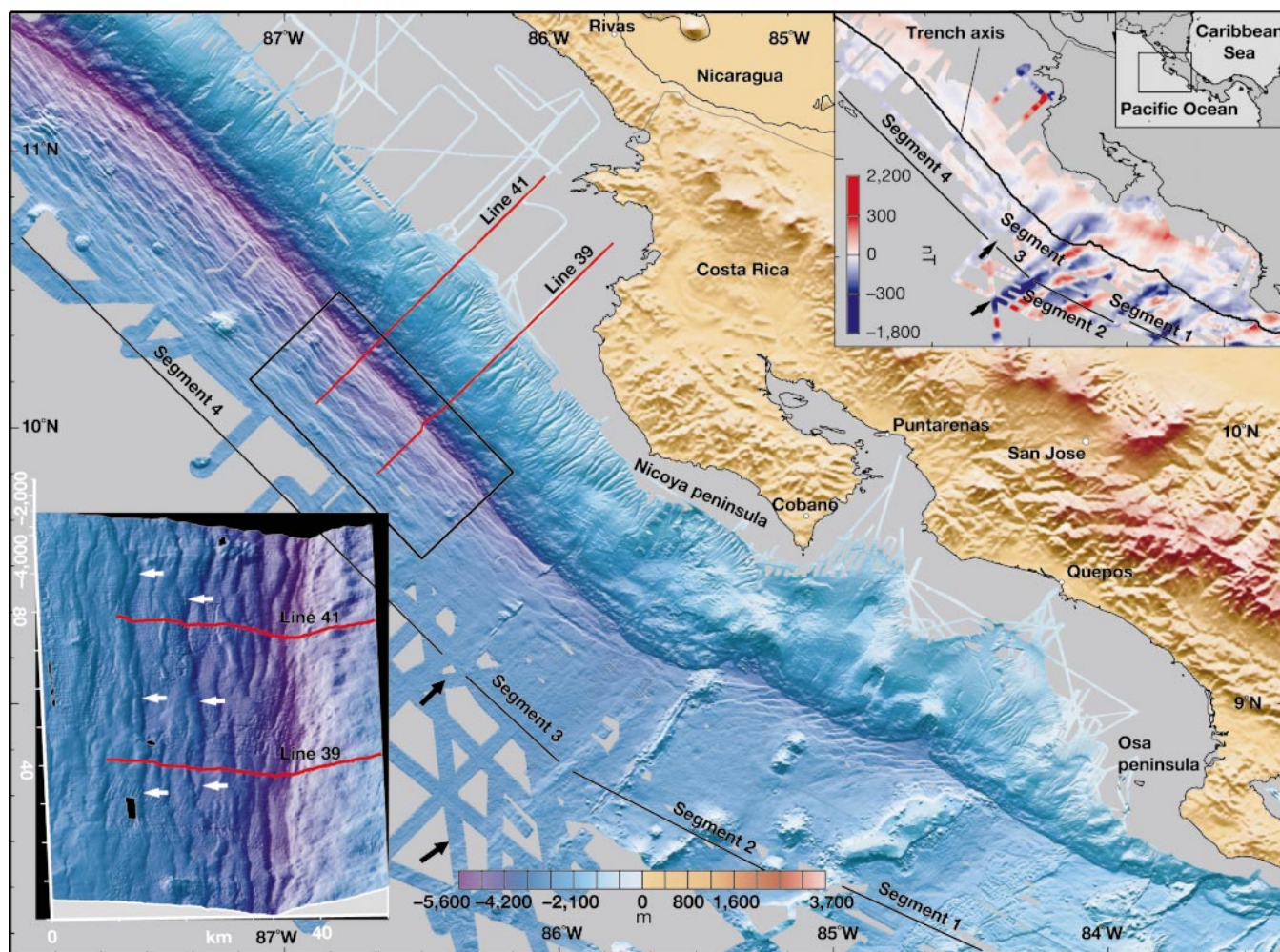


Figure 1 Colour-coded bathymetry and elevation map along ~ 600 km of the Middle America trench. Bathymetric soundings have been edited before gridding into a ~ 100 m grid spacing. Land topography is from a 30-arcsecond data base from US geological survey. Red lines are tracks from seismic profiles in Fig. 2. The data display the deformation of the ocean plate related to bending before subduction. Abrupt boundaries separate four distinctly faulted segments of the ocean plate^{17,47} (black arrows mark segment boundaries). Segment 1 is off southern Costa Rica, where the 12–20-km-thick crust of the Cocos ridge bends little, and little to no fault relief develops near the trench. The large fault scarps perpendicular to the trench predate sediment infill, indicating that they formed ~ 14 Myr ago when the Galapagos hotspot affected this area and are not related to subduction. Segment 2 has ~ 7 -km-thick crust and magnetic anomalies striking oblique to the trench. Here pervasive faulting with a strike similar to the original seafloor fabric develops within ~ 25 km of the trench axis. Segment 3 has ~ 6 -km-thick crust, a smooth morphology and magnetic anomalies striking perpendicular to the trench axis. In segment 3, a few small-offset faults develop, roughly parallel to the trench and at

high angles with the inherited tectonic fabric. Here the plate bends less than at contiguous segments and little deepening occurs towards the trench axis. A small linear scarp in the ocean plate separates segments 3 and 4. Lithosphere to the SE (segments 1 to 3) formed at the Galapagos spreading centre (see upper inset), whereas lithosphere to the NW (segment 4) formed at the East Pacific Rise. Although the lithosphere of segments 2 to 4 formed at similar spreading rates, major changes in plate bending and faulting occur. Segment 4 has 5–6-km-thick crust, magnetic anomalies striking parallel to the trench and pervasive normal faulting. The faulted region is the widest in this region, increasing towards the north from ~ 25 to ~ 60 km, while individual fault throws increase from ~ 100 to ~ 500 m. The two sets of white arrows in the lower inset mark seafloor offsets of two ~ 50 -km-long faults; these faults are imaged as clear deep penetrating faults in the seismic sections (Fig. 2). Lower inset shows a perspective close-up view of the bathymetry projected along the trench (100×50 km). Area is indicated by the box in regional figure and view is from SE. Upper inset shows a magnetic map of the area¹⁷.

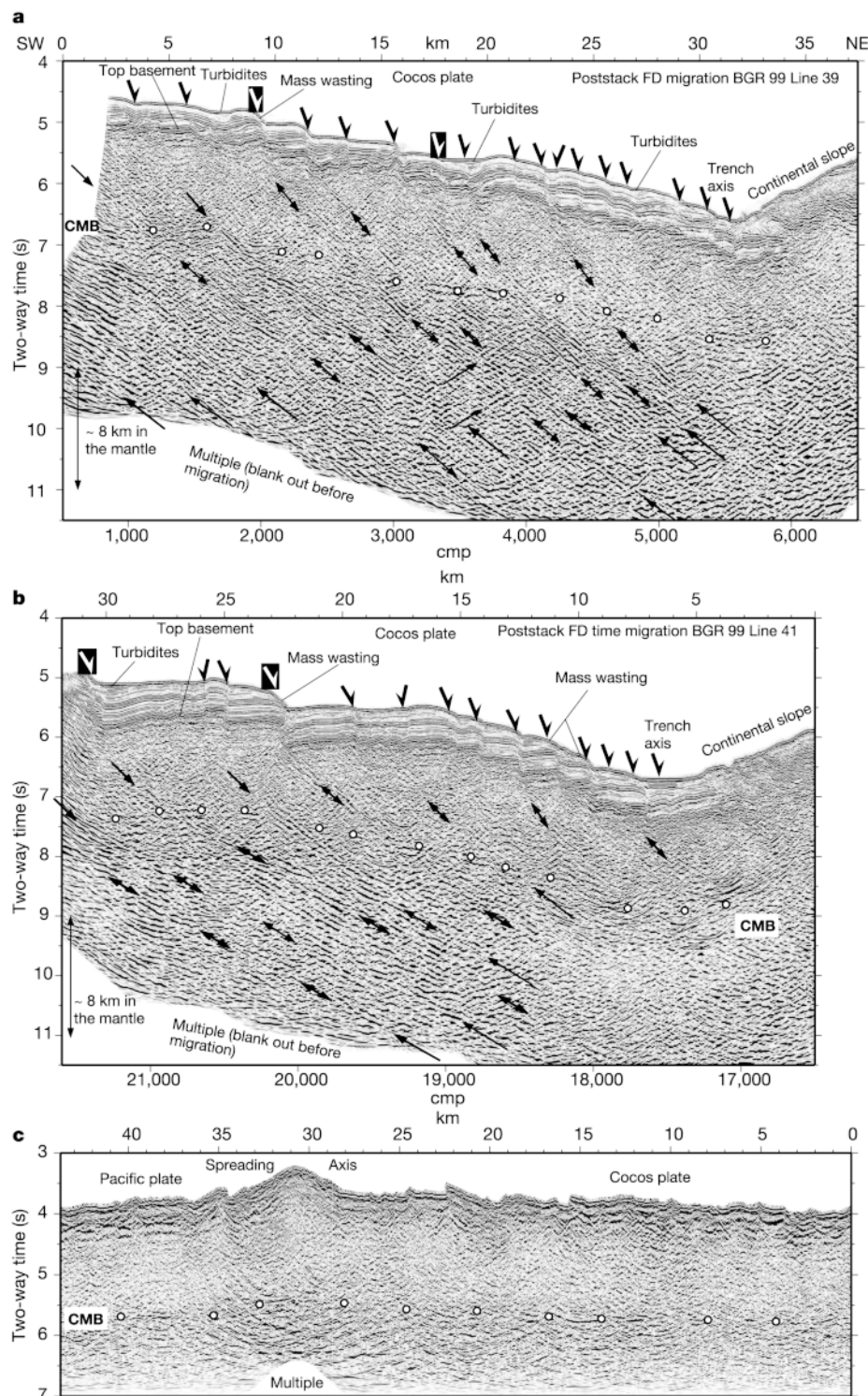


Figure 2 Poststack finite-difference time migration of various lines. **a**, Line 39 and **b**, line 41 of cruise BGR99 and **c**, line 1331 of Ewing cruise 9503. Lines 39 and 41 show the structure of the Cocos plate across the ocean trench slope. The crust–mantle boundary (CMB) is defined by a series of gently dipping reflections, delineated by white filled circles, at ~ 1.7 s two-way time (~ 5 km) beneath the top of the igneous basement. A pervasive fabric traverses the mantle: sometimes this fabric appears to continue up into reflections that cross the crust projecting into top-basement and seafloor offsets. Several reflections align forming individual features—indicated by the black arrows—that cross much of the lithosphere. We propose that these reflections represent normal faults which slip during plate bending. The reflections cut to ~ 18 – 20 km across the crust

into the mantle where they are masked by the acoustic energy of the seafloor multiple. Arrows at the sea floor indicate faults used for Figs 3 and 4. Seafloor mass-wasting processes are evidenced by truncation of strata at many fault scarps and by turbidite infill of half-grabens. White arrows in black boxes indicate faults marked by arrows in the lower inset of Fig. 1. Line Ewing 9503-1331 shows the structure of the ocean plate formed at the East Pacific Rise at latitude 16° N (ref. 27). The crust–mantle boundary is defined by discontinuous reflections at ~ 1.8 s two-way time, the crust and upper mantle are featureless. The ‘smiles’ underneath the spreading centre axis are created by overmigration of side-scattered energy in the rough axial topography. FD, finite difference.

Trenchward-dipping reflections across the crust and mantle project updip to the largest offsets in top basement and sea floor, indicating that these represent reflections from fault planes. Fault offsets at the sea floor are <200 m—that generate <50 ms travel-time offsets at the Moho, which explains the lack of clear offsets across the crust–mantle boundary reflections. Although these faults appear to form by reactivation of faults created at the palaeosspreading centre, such a deeply penetrating tectonic fabric could not have originated during crustal accretion at the palaeosspreading centre. Modern MCS data collected over sea floor created at fast-spreading mid-ocean ridges (MOR) demonstrates that the reflective mantle fabric observed at the MAT was not formed by accretion/deformation processes at the fast-spreading East Pacific Rise (EPR).

Since the first modern MCS reflection survey was collected across the EPR at 9°–13° N, where the Cocos plate subducting beneath Middle America is formed, seismic images have shown a “reflection-free” mantle underlying the Moho reflection²⁵. More recent MCS data, collected along segments of the EPR, comparable in quality to our near-trench data (for example, at 14°–18° S (ref. 26); and at 15.5°–16.5° N (ref. 27)), support earlier results (Fig. 2c). Seismic images of old crust formed at fast-spreading rates have regions of reflectivity limited to the crust and no evidence for crustal-scale brittle faulting²⁴. Furthermore, theoretical modelling indicates that at the axis of fast-spreading MORs, the brittle layer overlying the persistent magma lens is only a few kilometres thick^{28,29}. At the MAT, the strength of the mantle-dipping reflections is comparable to the Moho reflection (Fig. 2a, b), indicating that a similar fabric at the EPR would have been detected with seismic methods had it also existed there (Fig. 2c).

Further evidence against the origin of the reflective mantle fabric by ductile deformation at fast-spreading MORs comes from studies of ophiolite exposures. The ductile foliation of exposed mantle peridotites lies at a low angle to the Moho³⁰, a much shallower angle than the ~45° observed in the mantle reflections. These obser-

variations are also consistent with numerical models predicting that the mantle deformation fabric would lie semi-parallel to the Moho boundary³¹.

Finally, data collected in the Cocos plate in the lithosphere halfway between the EPR crest and the Nicaraguan trench shows that reflectivity in the mantle is not produced during the ageing of the ocean plate as it drifts towards a subduction zone³². Dipping reflections observed in old lithosphere formed at other fast-spreading MORs are constricted to the lower crust²⁴. Even in crust formed at slow-spreading rates, most seismically imaged normal faults are restricted to crustal levels³³, penetrating (rarely) at most a few kilometres across the Moho—again with a reflectivity pattern completely unlike that seen at the MAT.

At the MAT, dipping reflections appear to be images of deeply propagating faults formed by reactivation of shallow crustal faults created at the spreading centre. At fast-spreading MORs, roughly half the faults dip inward and half dip outward³⁴. As lithosphere bends at trenches, faults can penetrate deeply into a 30–40-km-thick cold and brittle layer. During initial bending near the trench both inward and outward faults may be reactivated²³, which could explain the few and indistinct seaward-dipping reflections in the mantle. However, the seismic images and fault scarps in the bathymetry show that landward-dipping faults prevail both in the seafloor fabric of this particular ocean trench slope and in the reflective fabric of its underlying lithospheric mantle (Fig. 2a, b).

Offsets of individual faults measured at the sea floor and the top of the igneous crust show consistent differences: the top basement fault offsets are in most cases larger, with the difference increasing approaching the trench (Fig. 3). These differences, and the truncation of strata at the sea floor near fault scarps (Fig. 2), indicate that fault offset at the sea floor is often obscured by mass wasting, which tends to decrease surface fault relief. Indeed, small wedges of turbidites are observed at the foot of some fault scarps (Fig. 2). This may be the reason why a study at a different area²² based upon

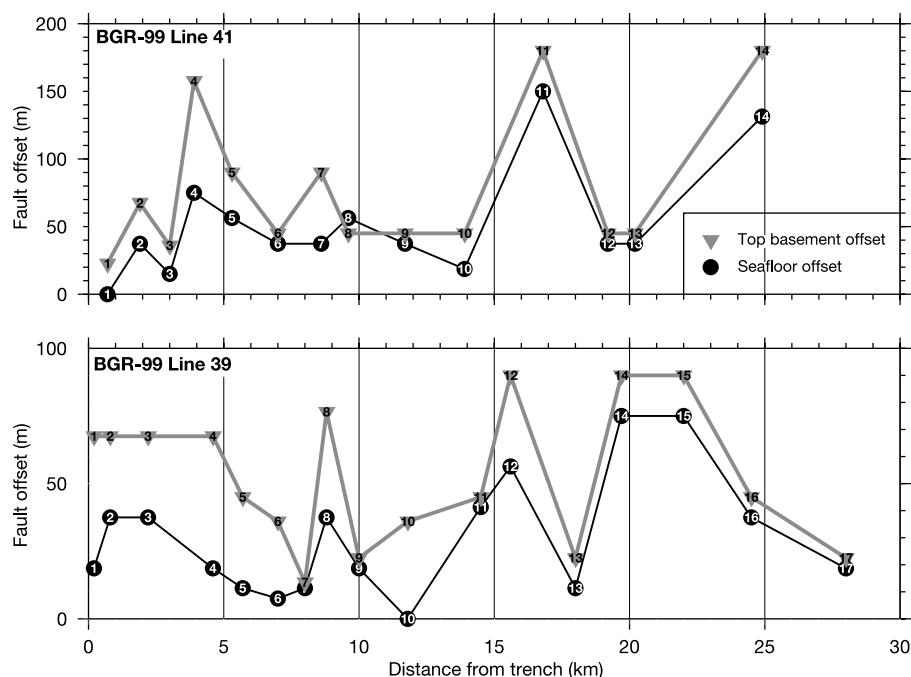


Figure 3 Fault offsets measured at the sea floor and top of igneous basement for faults marked on Fig. 2, plotted versus distance to the trench axis. The difference in fault offset at the sea floor and top basement is readily explained as a by-product of active mass-wasting processes that decrease the offsets of seafloor scarps. In this case, near-trench faults with similar seafloor and top basement offsets (for example, faults 7 and 9 on line 39

and faults 5, 7 and 8 on line 41) indicate recently formed faults, which typically have small offsets. Faults showing large differences in offsets (for example, faults 1–6 on line 39, and 3, 4 and 6 on line 41) would be faults formed away from the trench axis which continued to slip and experience surface mass-wasting as they approached the trench axis.

only bathymetric data suggested that active faulting stops in the middle of the trench slope. Here, the increase in the difference in fault offset between sea floor and top basement approaching the trench indicates that faulting remains active across the entire trench slope. Both the average fault offset and the number of faults increase towards the trench (Figs 3 and 4), implying both continued fault slip and fault formation in this region.

The observations can be explained with a model of sequential fault generation during progressive plate bending. First, widely spaced faults are reactivated during initial plate bending at the outer rise. Next, the segments between faults are cut by new faults as the plate approaches the trench axis and bending continues. Recent faults formed near the trench have similar fault offsets (see faults 6, 8 and 9 at line 41 and faults 7, 9 and 11 at line 39 in Fig. 3) at the sea floor and top basement, whereas older faults have a larger offset at the top basement than at the sea floor (see faults, 4, 5 and 7 on line 41 and faults 1–6 and 8 on line 39 in Fig. 3).

Where bend-faulting occurs, extension remains active across the entire oceanward trench slope, providing an environment where fluids can be persistently pumped down into the chemically reactive lithospheric mantle and crust. Offshore of Nicaragua, the thermal structure of the ~23-Myr-old plate implies that the brittle–ductile transition (~600 °C) is ~35 km deep; earthquakes may rupture to the mantle depths indicated by fault reflections. This scenario may promote pumping of water³⁵ across the crust into the mantle. At temperatures <500–600 °C mantle peridotite is unstable with respect to water and readily ‘corrodes’ to serpentine. Serpentinization leads to important physical and chemical changes in the subducting lithosphere. It involves a large volume increase—complete serpentinization of a peridotite decreases its density by ~40%—promoting self-sealing closure of any open fluid conduits along faults. However, persistent brittle faulting across the entire trench slope may effectively reopen and recreate pathways for intensive fluid transport into the lithosphere.

The seismic data images a series of discrete reflections across the ocean crust and a more pervasive tectonic fabric beneath the crust–mantle boundary (CMB) (Figs 2a, b). Some of the most conspicuous dipping reflections in the mantle do not extend up into crustal reflections, while in other cases bright mantle reflections project updip into indistinct crustal reflections. We suggest that serpentinization (and perhaps also free fluids) produces the bright mantle reflectivity; the small offsets along observed faults would not be enough to create an anisotropic fabric detectable by seismic

reflection methods. The more pervasive reflection fabric of the mantle does not necessarily indicate that the deformation between crust and mantle is decoupled but rather that tectonic structures in the mantle are easier to detect because mantle peridotites hydrate more easily and extensively than basaltic ocean crust.

The spacing in between packages of reflections defining faults in the mantle is somewhat shorter than between fault scarps at the sea floor (Fig. 2). A possible explanation is that at mantle depths where velocities are close to 8 km s⁻¹, the dominant frequencies of the returned signal (10–15 Hz) are lower than in the signal coming from the sediment section (45–80 Hz), so the deep portion of the faults may image as a broader area. Also, the Fresnel zone, that is, the dimensions of the lateral area sampled by the seismic wavefield, is broader for long travel times and thus the image of mantle reflectors is not as sharp as for shallow faults because the seismics are mapping a three-dimensional (3D) structure onto a two-dimensional cross-section. Furthermore, fault zones usually broaden as they deepen, so that the faults we observe may not be simple planes but rather packages of fractures. At the sea floor, arrays of faults arrange in longer fault segments (Fig. 1, lower inset). If faults broaden into the crust and mantle, fracturing may affect a larger volume at depth than at the sea surface.

The observed 50-km-wide region of pervasive faulting offshore of Nicaragua between the onset of faulting and the trench axis (Fig. 1) implies, at current convergence rates, that extension and water percolation have occurred there for at least 0.7 Myr. This time span is similar to the time that ocean lithosphere faults at slow-spreading MORs, but offshore of Nicaragua the water can penetrate at least four times deeper into the mantle before reaching the 600 °C isotherm. The incoming ocean lithosphere along the circum-Pacific subduction zones was almost entirely formed at intermediate- to fast-spreading MORs where axial faulting and associated vigorous hydrothermal circulation is believed to be shallower (~2–3 km depth) than typical at slow-spreading centres (~8–10 km depth). Thus bend-faulting, in addition to hydrating the mantle, may lead to pervasive pre-subduction hydration of crust formed at fast-spreading MORs, perhaps comparable to the crustal hydration experienced at slow-spreading MORs.

Faulting and intraslab seismicity

Faults in the seismic data dip ~45° and cut to at least 20 km into the plate (Fig. 2). Similarly, nodal planes of outer rise earthquakes elsewhere dip ~40–45° (refs 15, 36). More controversial is the proposition that earthquake rupture depths extend to ≥20 km (for example, ref. 36). The depth of penetration of faulting seismically imaged here supports the hypothesis that earthquakes can cut down at least 20 km beneath the sea floor.

The combination of swath bathymetry and seismic data provides insights into the 3D structure and dimensions of bend-faults. Some faults imaged on both seismic lines project updip to the same fault scarp at the sea floor. Two examples of deep penetrating faults are marked in Fig. 2 with white arrows. These faults cut the sea floor at the scarps marked with white arrows in the lower inset of Fig. 1—scarps which extend for 50 km or more. Thus individual fault planes can be ~50 km long along strike by ~20 km deep into the plate—dimensions similar to, or somewhat larger, than the size of the rupture area of individual intermediate depth earthquakes⁵.

The reflectivity of the small-offset, deep-penetrating faults imaged in the seismic data can be best explained by water percolation and mineral alteration along the fault planes. Support for robust hydrothermal circulation associated with bend-faulting comes from the low heatflow values measured at the trench off Costa Rica³⁷ and elsewhere³⁸. Thus serpentinized bend-faults may become regions where increased relative concentrations of hydrous minerals promote rupture of a fault several times during its descent. Crustal sections should generally dehydrate at shallower depths than mantle sections, in part because of the lower pressure/temperature

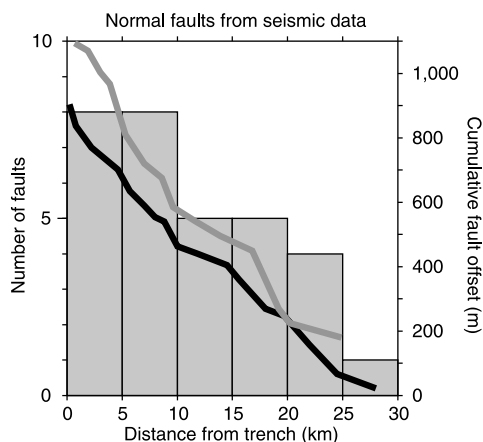


Figure 4 Histogram of fault density versus distance to the trench axis. The number of faults increases towards the axis, indicating that new faults continue to form as the plate approaches the trench axis. Lines indicate the cumulative fault offset measured at top basement for line BGR99-39 (black line) and BGR99-41 (grey line). There is a tendency to larger fault offset towards the trench.

stability field of hydrated basalt than serpentine, but even more because the crust of the subducting plate reheats before the mantle. Likewise, serpentinites in the coolest slab-centre portions of faults may dehydrate deeper within the subduction zone³⁹. In North Chile, intraslab earthquakes occur in the ocean crust at 100–150 km depth, and in the slab mantle at depths >150 km (ref. 40). Similarly, double seismic zones elsewhere (New Zealand, Tonga, New Britain, Mariana, NE Japan, Kuril, Kamchatka, Aleutian, Alaska, Gorda Plate, Mexico, northern Chile, central Chile^{7,41,42}) have a separation between upper and lower seismogenic planes of 15–30 km, with the lower plane reaching deeper into the subduction zone.

The slab fluid budget, recycling and arc magmatism

Faulting along trenches occurs along most of the world subduction zones as indicated by local swath bathymetry data^{21–23} and outer rise earthquakes^{15,16}. To evaluate the importance of hydration related to bend-faulting on the global budget of water in subduction zones, we estimate the potential volume of water contained in serpentinized mantle offshore of Nicaragua.

For the calculation we assume that serpentinization occurs along the large faults spaced ~2–3 km in the 25 km closest to the trench (Fig. 4), with vigorous hydration beneath the trench slope occurring for 0.3 Myr at the current convergence rate of ~80 km Myr⁻¹. The diffusion speed of water in unfractured serpentine is ~1 km Myr^{-1/2} (ref. 43), implying a maximum potential for ~30% serpentinization of the uppermost mantle if water is always available along faults and diffuses away from the faults through a 'rind' of serpentinized mantle. Consider that the serpentinization is 3–30% at Moho depths and decreases linearly to 0% at the ~30 km or 600 °C thermal depth limit to serpentinization in ~23-Myr-old lithosphere (the depth where brittle faulting may stop). Chemically bound water forms 13 wt% of serpentine, implying that this 30-km-thick mantle section with 1.5–15% average serpentinization would hold the chemically bound equivalent of a column of 0.17–1.7 km of water per unit length of trench. Previous estimates of 0.14 km (ref. 12)—1 km (ref. 13) for the non-sediment water supplied to subduction zones have generally focused on the water contained in the ocean crust. The above crude estimate for Nicaragua suggests that serpentinized mantle may have the potential to subduct as much chemically bound water as does the oceanic crust.

Additionally, serpentinization may provide a geochemical pathway for recycling water-soluble trace elements into both the sub-arc region of flux-melting and the deeper mantle. Furthermore, sediments are likely to lose more than two-thirds of their chemically bound water by ~50 km depth, and amphibolitized crust (and sediments) will dehydrate almost completely after subducting to ~120 km in typical subduction zone environments³⁹. But antigorite serpentine is one of the most stable hydrous phases⁴⁴, so that serpentinized lithosphere will progressively dehydrate until the ~200 km depth of the antigorite-phase A hydrous phase transformation¹⁰, at which point dehydration greatly slows or stops.

This implies that antigorite will be the primary source of hydrous fluids from the dewatering slab between 120–200 km depths, potentially providing a key fluid-component to arc-magmas. As these reducing high-temperature fluids migrate upward through the overlying crust and sediments they may be able to scavenge Pb, Be, and similar-behaviour elements from the slab crust, and sediments through which they migrate³⁹, leading to the preferential hydrous 'leaching' of Pb with respect to Ce that is thought to be linked to the formation of the 'HIMU' mantle source component⁴⁵. Serpentine that transforms to phase A may also recycle a geochemically significant amount of rare gases into the mantle (sea water contains 1%, and 3% of the exosphere's Ar, and Xe, respectively), providing an effective mechanism for this previously hypothesized geochemical process⁴⁶.

Thus our observations demonstrate that the oceanic lithosphere undergoes widespread deformation during bending at the MAT.

Active faulting across the entire ocean trench slope penetrates at least 20 km into the plate. We infer that faulting provides a mechanism for pervasive infiltration of water into the crust and mantle. This leads to a profound chemical transformation of the lithosphere during which it acquires a series of characteristics that help to explain the generation of intermediate-depth earthquakes and arc magmatism. Although the amount of bend-faulting may change within a few hundred kilometres of distance along a trench, this change is usually linked to locally thickened crust. Oceanic lithosphere rapidly attains much of its thickness during the first ~25 Myr after formation, whereas for older ages thickening slows down considerably. Assuming that the depth of fault penetration is mainly controlled by bending stresses and rheology, we anticipate that similar bend-faulting related processes occur along most subduction zones. □

Received 1 October 2002; accepted 23 July 2003; doi:10.1038/nature01961.

1. Savage, J. C. The mechanics of deep-focus faulting. *Tectonophysics* **8**, 115–127 (1969).
2. Frohlich, C. The nature of deep-focus earthquakes. *Annu. Rev. Earth Planet. Sci.* **17**, 227–254 (1989).
3. Raleigh, C. B. & Paterson, M. S. Experimental deformation of serpentinite and its tectonic implications. *J. Geophys. Res.* **70**, 3965–3985 (1965).
4. Kirby, S., Engdahl, E. R. & Denlinger, R. in *Subduction: Top to Bottom* (eds Bebout, G. E., Scholl, D., Kirby, S. H. & Platt, J. P.) 195–214 (Geophysical Monograph 96, American Geophysical Union, Washington, 1996).
5. Tibi, R., Bock, G. & Estabrook, C. H. Seismic body wave constraint on mechanisms of intermediate-depth earthquakes. *J. Geophys. Res.* **107**, 101029/2001JB000361 (2002).
6. Meade, C. & Jeanloz, R. Deep focus earthquakes and recycling of water into the Earth's mantle. *Science* **252**, 68–72 (1991).
7. Peacock, S. Are the lower planes of double seismic zones caused by serpentine dyhydration in subducting oceanic mantle? *Geology* **29**, 299–302 (2001).
8. Jiao, W., Silver, P. G., Fei, Y. & Prewitt, C. T. Do intermediate- and deep-focus earthquakes occur on preexisting weak zones? An examination of the Tonga subduction zone. *J. Geophys. Res.* **105**, 28125–28138 (2000).
9. Gill, J. *Orogenic Andesites and Plate Tectonics* (Springer, New, 1981).
10. Ulmer, P. & Trommsdorff, V. Serpentine stability to mantle depths and subduction-related magmatism. *Science* **268**, 858–861 (1995).
11. Ernst, W. G. Hornblende. The continent maker—Evolution of H₂O during circum-Pacific subduction versus continental collision. *Geology* **27**, 675–678 (1999).
12. Peacock, S. Fluid processes in subduction zones. *Science* **248**, 329–337 (1990).
13. Moore, J. C. & Vrolijk, P. Fluids in accretionary prisms. *Rev. Geophys.* **30**, 113–135 (1992).
14. Seno, T. & Yamanaka, Y. in *Subduction: Top to Bottom* (eds Bebout, G. E., Scholl, D., Kirby, S. H. & Platt, J. P.) 347–355 (Geophysical Monograph 96, American Geophysical Union, Washington, 1996).
15. Christensen, D. H. & Ruff, L. Seismic coupling and outer rise earthquakes. *J. Geophys. Res.* **93**, 13421–13444 (1988).
16. Hasegawa, A., Horiuchi, S. & Umino, N. Seismic structure of the northeastern Japan convergent margin: A synthesis. *J. Geophys. Res.* **99**, 22295–22311 (1994).
17. Barchhausen, G. A., Ranero, C. R., von Huene, R., Cande, S. & Roeser, H. Revised tectonic boundaries in the Cocos Plate off Costa Rica: Implications for the segmentation of the convergent margin and for plate tectonic models. *J. Geophys. Res.* **106**, 19207–19220 (2001).
18. Protti, M., Guendel, F. & McNally, K. Correlation between the age of the subducting Cocos plate and the geometry of the Wadati-Benioff zone under Nicaragua and Costa Rica. *Geol. Soc. Am. (Spec. Pap.)* **295**, 309–326 (1995).
19. Li, C. Forearc structures and tectonics in the Southern Peru–Northern Chile continental margin. *Mar. Geophys. Res.* **17**, 97–113 (1995).
20. von Huene, R. et al. Tectonic control of the subducting Juan Fernández Ridge on the Andean margin near Valparaíso, Chile. *Tectonics* **16**, 474–488 (1997).
21. von Huene, R. & Ranero, C. R. Subduction erosion and basal friction along the sediment starved convergent margin off Antofagasta Chile. *J. Geophys. Res.* **108**(2079), 101029/2001JB001569 (2003).
22. Kobayashi, K., Nakanishi, M., Tamaki, K. & Ogawa, Y. Outer slope faulting associated with the western Kuril and Japan trenches. *Geophys. J. Int.* **134**, 356–372 (1998).
23. Masson, D. G. Fault patterns at outer trench walls. *Mar. Geophys. Res.* **13**, 209–225 (1991).
24. Ranero, C. R., Reston, T. J., Belykh, I. & Gnibidenko, H. Reflective oceanic crust formed at a fast-spreading center in the Pacific. *Geology* **25**, 499–502 (1997).
25. Barth, G. A. & Mutter, J. C. Variability in oceanic crustal thickness and structure: Multichannel seismic reflection results from the northern East Pacific Rise. *J. Geophys. Res.* **101**, 17951–17975 (1996).
26. Detrick, R. S. et al. Seismic structure of the southern East Pacific Rise. *Science* **259**, 499–503 (1993).
27. Carbotte, S. M., Solomon, A. & Ponce-Correa, G. Evaluation of morphological indicators of magma supply and segmentation from a seismic reflection study of the East Pacific Rise 15°30'–17°N. *J. Geophys. Res.* **105**, 2737–2759 (2000).
28. Phipps Morgan, J., Harding, A., Orcutt, J., Kent, G. & Chen, Y. J. in *Magmatic Systems* (ed. Ryan, M. P.) 139–178 (Academic, 1994).
29. Morton, J. & Sleep, N. A mid-ocean ridge thermal model: constraints on the volume of axial hydrothermal flux. *J. Geophys. Res.* **90**, 11345–11353 (1985).
30. Nicolas, A. *Structures of Ophiolites and Dynamics of Oceanic Lithosphere* (Kluwer Academic, Dordrecht, 1989).
31. Phipps Morgan, J. & Chen, Y. J. The genesis of oceanic crust: Magma injection, hydrothermal circulation, and crustal flow. *J. Geophys. Res.* **98**, 6283–6297 (1993).

32. Hallenborg, E., Harding, A. J., Kent, G. M. & Wilson, D. S. Seismic structure of 15 Ma oceanic crust formed at an ultra-fast spreading East Pacific Rise: evidence for kilometer-scale fracturing from dipping reflectors. *J. Geophys. Res.* (in the press).
 33. Ranero, C. R., Banda, E. & Buhl, P. The crustal structure of the Canary Basin: Accretion processes at slow spreading centers. *J. Geophys. Res.* **102**, 10185–10201 (1997).
 34. Carbotte, S. M. & Macdonald, K. C. Comparison of the seafloor tectonic fabric at intermediate, fast, and super fast spreading ridges: Influence on spreading rate, plate motions, and ridge segmentation on fault patterns. *J. Geophys. Res.* **99**, 13609–13631 (1994).
 35. Sibson, R. H. Controls on low-stress hydrofracturing dilatancy in thrust, wrench and normal fault terrains. *Nature* **289**, 665–667 (1981).
 36. Lynnes, C. S. & Lay, T. Source process of the great 1977 Sumba earthquake. *J. Geophys. Res.* **93**, 13407–13420 (1988).
 37. Langseth, M. G. & Silver, E. The Nicoya convergent margin—a region of exceptionally low heat flow. *Geophys. Res. Lett.* **23**, 891–894 (1996).
 38. Yamano, M. & Uyeda, S. Heat-floor studies in the Peru trench subduction zone. *Proc. ODP Sci. Res.* **112**, 653–661 (1990).
 39. Ruepke, L. H., Phipps Morgan, J., Hort, M. & Connolly, J. A. D. Are the regional variations in Central American arc lavas due to differing basaltic versus peridotitic slab sources of fluids? *Geology* **30**, 1035–1038 (2002).
 40. Bock, G., Schurr, B. & Asch, G. High-resolution image of oceanic Moho in the subducting Nazca plate from P-S converted waves. *Geophys. Res. Lett.* **27**, 3929–3932 (2000).
 41. Cassidy, J. F. & Waldhauser, F. Evidence for both crustal and mantle earthquakes in the subducting Juan de Fuca plate. *J. Geophys. Res.* **108**, 10101029/2002GL015511 (2003).
 42. Hacker, B. R., Peacock, S. M., Abers, G. A. & Holloway, S. D. Subduction factory 2. Are intermediate-depth earthquakes in subducting slabs linked to metamorphic dehydration reactions? *J. Geophys. Res.* **108**, 20101029/2001JB001129 (2003).
 43. Macdonald, H. & Fyfe, W. S. Rate of serpentinization in seafloor environments. *Tectonophysics* **116**, 123–135 (1985).
 44. Schmidt, M. W. & Poli, S. Experimentally based water budgets for dehydrating slabs and consequences for arc magma generation. *Earth Planet. Sci.* **163**, 361–379 (1998).
 45. Chauval, C., Hofmann, A. W. & Vidal, P. HIMU-EM: The French-Polynesian connection. *Earth Planet. Sci. Lett.* **110**, 99–119 (1992).
 46. Porcelli, D. & Wasserburg, G. J. Mass transfer of helium, neon, argon, and xenon through a steady-state upper mantle. *Geochim. Cosmochim. Acta* **59**, 4921–4937 (1995).
 47. Ranero, C. R. & Von Huene, R. Subduction erosion along the Middle America convergent margin. *Nature* **404**, 748–752 (2000).
- Acknowledgements** The bathymetry data were collected during R/V *Sonne* cruises 76, 81, 107, 144, 150 and 163, R/V *Meteor* cruise 54 and R/V *M. Ewing* cruises 0005 (Chief Scientist K. McIntosh) and 0104 (Chief Scientist A. Fisher). The seismic reflection data were collected during BGR99 cruise aboard M/V *Professor Polshkov*. We thank the scientific parties for their efforts and officers and crews for their technical and logistical support. R/V *Sonne* cruises were funded by Deutsche BMBF, the R/V *Meteor* cruise by the DFG and the R/V *M. Ewing* cruises by the NSF-USA. This work is a contribution of ‘SFB574 Volatiles and Fluids in Subduction Zones’ from the University of Kiel.
- Competing interests statement** The authors declare that they have no competing financial interests.
- Correspondence** and requests for materials should be addressed to C.R.R. (cranero@geomar.de).

A test of general relativity using radio links with the Cassini spacecraft

B. Bertotti¹, L. Iess² & P. Tortora³

¹Dipartimento di Fisica Nucleare e Teorica, Università di Pavia, Via U. Bassi 6, I-27100, Pavia, Italy

²Dipartimento di Ingegneria Aerospaziale ed Astronautica, Università di Roma "La Sapienza", Via Eudossiana 16, I-00184, Roma, Italy

³Il Facoltà di Ingegneria, Università di Bologna, Via Fontanelle 40, I-47100, Forlì, Italy

According to general relativity, photons are deflected and delayed by the curvature of space-time produced by any mass^{1–3}. The bending and delay are proportional to $\gamma + 1$, where the parameter γ is unity in general relativity but zero in the Newtonian model of gravity. The quantity $\gamma - 1$ measures the degree to which gravity is not a purely geometric effect and is affected by other fields; such fields may have strongly influenced the early Universe, but would have now weakened so as to produce tiny—but still detectable—effects. Several experiments have confirmed to an accuracy of $\sim 0.1\%$ the predictions for the deflection^{4,5} and delay⁶ of photons produced by the Sun. Here we report a measurement of the frequency shift of radio photons to and from the Cassini spacecraft as they passed near the Sun. Our result, $\gamma = 1 + (2.1 \pm 2.3) \times 10^{-5}$, agrees with the predictions of standard general relativity with a sensitivity that approaches the level at which, theoretically, deviations are expected in some cosmological models^{7,8}.

Testing theories of gravity in the Solar System and with binary pulsars has been pursued for a long time^{1,2}, yet general relativity has survived whereas most of its alternatives have been disproved. In particular, the other main test—the anomalous advance of the pericentre of an orbiting body, such as Mercury around the Sun—has been found in agreement with Einstein's prediction, with a similar accuracy $\sim 0.1\%$. In the past 20 yr there has been no appreciable improvement. With the Cassini mission, this barrier has now been largely overcome as far as γ is concerned, but no violations of general relativity have been detected.

The increase Δt produced by the gravitational field of the Sun (with mass M_S and radius R_S) in the time taken for light to travel the

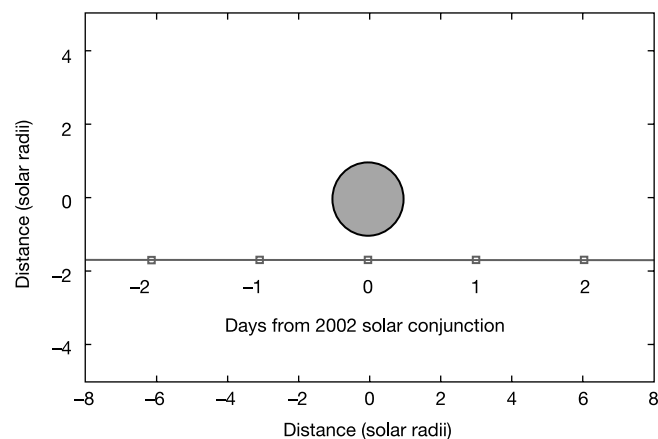


Figure 1 Geometry of the 2002 Cassini solar conjunction. The graph shows Cassini's motion in the sky relative to the Sun, as a function of days from the 2002 solar conjunction; coordinates are in solar radii. The conjunction—at which the spacecraft (at a geocentric distance of 8.43 au), the Sun and the Earth were almost aligned, in this order—occurred on 21 June 2002, with a minimum impact parameter $b_{\min} = 1.6 R_S$, and no occultation.

round trip between the ground antenna and the spacecraft, at distances r_1 and r_2 respectively from the Sun, is¹:

$$\Delta t = 2(1 + \gamma) \frac{GM_S}{c^3} \ln \left(\frac{4r_1 r_2}{b^2} \right) \quad (1)$$

where G is the gravitational constant, b ($\ll r_1, r_2$) the impact parameter and c the velocity of light. The motion of the spacecraft and Earth produces a change in b and Δt , equivalent to a change in distance, and hence a change in relative radial velocity. The corresponding fractional frequency ($y_{\text{gr}} = \Delta\nu/\nu$) shift for a two-way radio signal is⁹:

$$y_{\text{gr}} = \frac{d\Delta t}{dt} = -2(1 + \gamma) \frac{GM_S}{c^3 b} \frac{db}{dt} = -(1 \times 10^{-5} \text{ s})(1 + \gamma) \frac{1}{b} \frac{db}{dt} \quad (2)$$

For a spacecraft much farther away from the Sun than the Earth, db/dt is not very different from the Earth's velocity $v_E = 30 \text{ km s}^{-1}$. In the Cassini solar conjunction the peak value of y_{gr} is 6×10^{-10} . The Cassini experiment, exploiting the new observable y_{gr} (refs 9, 10), was carried out between 6 June to 7 July 2002, when the spacecraft was on its way to Saturn, around the time of a solar conjunction (Fig. 1). The gravitational signal and the tracking passes that provided useful data are shown in Fig. 2.

The main reason why the Doppler method has not been applied before is the overwhelming noise contribution due to the solar corona. The Cassini mission has overcome this hindrance with: (1) high-frequency carrier waves in the Ka-band, in addition to the X-band for standard operation; and (2) a multi-frequency link in which three different phases are measured at the ground station^{11,12}. Two carriers at 7,175 MHz (X-band) and 34,316 MHz (Ka-band) are transmitted from the ground; whereas, in addition to the downlink carriers at 8,425 MHz and 32,028 MHz locked on board to the X and the Ka signals respectively, a nearby Ka-band downlink carrier coherent with the X-band uplink is also transmitted back. This novel radio configuration uses dedicated and advanced instrumentation, both on board the spacecraft and at the ground antenna, and allows a full cancellation of the solar plasma noise (see Supplementary Fig. S1 for details)^{13–15}. The resulting measurement errors are four orders of magnitude smaller than the relativistic signal in equation (2).

The new ground station DSS25 at the NASA Deep Space Network complex in Goldstone, California, has performed admirably, par-

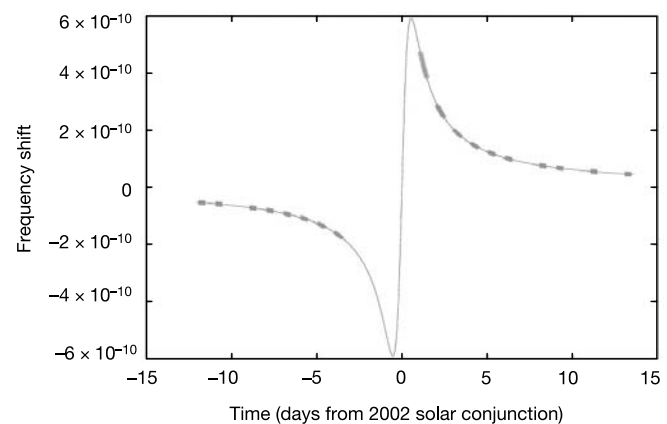


Figure 2 The gravitational signal. The two-way relativistic frequency shift y_{gr} due to the Sun and the available 18 passages, each lasting about 8 h, is shown. Unfortunately, no data could be acquired for three days just before conjunction owing to a failure of the ground transmitter; moreover, the tracking data acquired near closest approach were particularly noisy. A much larger plasma noise was detected in some passes after conjunction, and it was fully removed by the multi-link technique. Remarkably, during this time period, SOHO observations revealed large coronal mass ejections traversing the radio beam.

ticularly with respect to its Ka-band capabilities and special instrumentation designed to measure variations in path delay due to the wet troposphere^{13,15}. A Ka–Ka frequency translator (the crucial onboard instrument, provided by the Italian Space Agency) has assured very high phase coherence between the uplink and downlink signals.

The solar corona adds to each carrier very large frequency fluctuations, impairing the accuracy of conventional closed-loop receivers when the ray path is very close to the Sun (within about 10 solar radii); for this experiment, a wide-band, open-loop receiver has been used instead. A specially designed digital frequency estimator¹⁴ has been applied to the received signal (sampled at 1 kHz) to obtain the three required time series $\gamma_{XX}(t)$, $\gamma_{KK}(t)$, $\gamma_{XK}(t)$ of the reconstructed sky frequencies with a resolution of 1 s, where the subscripts X and K correspond to the X-band and Ka-band, respectively.

The final processing of the plasma-free, non-dispersive fractional frequency shift $\gamma_{nd}(t)$ requires the calibration of the dry and wet (water vapour) components of the Earth's troposphere. Its dry part can be determined from ground meteorological data and suitable modelling of elevation-dependent effects. In terms of fractional frequency shift, its magnitude can be as large as 10^{-12} . The wet component is much smaller ($\sim 3 \times 10^{-14}$, over timescales of 1,000–10,000 s) but much more difficult to model. Its contribution was measured at the ground station with especially designed radiometers that are able to measure the water vapour content along the line of sight. In this way the estimated total tropospheric contribution to the optical path variation could be subtracted from the signal, with a very significant improvement of the results.

An important contribution to the relative frequency shift is due to non-gravitational forces acting on the spacecraft. During the experiment the spacecraft was kept in a stable dynamical and thermal state, with its attitude controlled only by momentum wheels, and without turning on or off any subsystems or instruments. No unexpected buffeting or acceleration was detected throughout the experiment. Thanks to its large mass (5,200 kg) and careful design, Cassini turned out to be an extremely stable platform, subject only to small and steady non-gravitational accelerations owing to the solar radiation pressure and the anisotropic thermal emission of the spacecraft (largely produced by three on-board radioisotope thermoelectric generators (RTGs)). The

RTGs dissipate about 10 kW of constant thermal power, in large part isotropically radiated. The non-isotropic part of this thermal emission produces an acceleration, whose components are constant in the spacecraft reference frame. Deriving this acceleration from a model of the spacecraft is a difficult task; but its estimation from Doppler measurements, combined with attitude data, is routinely carried out for spacecraft navigation with good and consistent accuracies. The largest component (along the Earth–spacecraft direction, pushing towards the Earth) is about $3 \times 10^{-9} \text{ m s}^{-2}$, more than four orders of magnitude smaller than the peak value of the relativistic acceleration signal $\gamma_{gr}c(db/dt)/b \approx 3.6 \times 10^{-5} \text{ m s}^{-2}$ (corresponding to a velocity change $\gamma_{gr}c$ in the timescale $b/(db/dt)$). This component has been determined with a formal error of $\sim 3\%$. The other two components (orthogonal to the orbital plane and in the orbital plane) do not significantly affect the orbit and are smaller (1×10^{-10} and $4 \times 10^{-10} \text{ m s}^{-2}$, respectively)—they are determined much less accurately (respectively to 100% and 50%). A better estimation of these components was possible using data from another radio science experiment carried out by Cassini at solar opposition, from 26 November 2002 to 4 January 2003, under similar conditions.

As a consequence of the large distance from the Sun and the small area-to-mass ratio of the spacecraft, the solar radiation produces an acceleration of about an order of magnitude smaller than the RTGs thermal thrust¹⁰. (The incident solar power is just 175 W, as compared with 10 kW for the RTGs.) Modelling this acceleration at conjunction is simple, as the spacecraft is shaded by the large (4-m diameter) parabolic high-gain antenna. In general, the direction of the force depends on the spacecraft attitude, in particular the solar aspect angle between the axis of the antenna and the Sun. For Cassini, this angle changed by only two degrees during the 30-day experiment and almost vanished at conjunction. Therefore, although the force is almost radial, information concerning the spacecraft attitude and trajectory must be used for its determination in the orbital fit. The magnitude of the thrust depends on the thermo-optical properties of the dish, in particular the specular and diffuse reflectivity coefficients. However, as the spacecraft attitude in an inertial frame does not change much, the orbital fit cannot effectively discriminate between them; nonetheless, the estimated values are consistent with those obtained from previous tracking data. The $1/r^2$ variation of the force (about 2.3% of the average

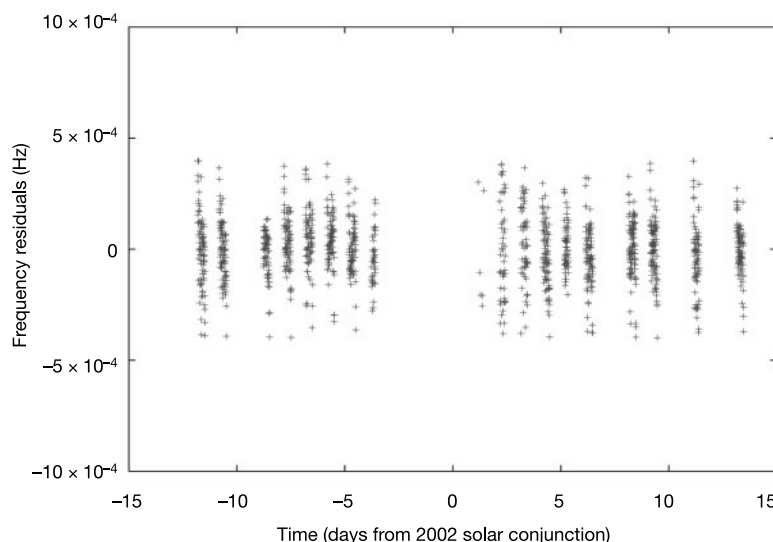


Figure 3 The Doppler residuals. Two-way frequency residuals, relative to an 8.4-GHz carrier, as a function of time (in days from solar conjunction) obtained from the final 9-parameter fit on the calibrated data. The r.m.s. value of the 1,094 data points

(compressed at a 300-s integration time) is $1.2 \times 10^{-4} \text{ Hz}$, corresponding to a one-way range rate of $2.2 \times 10^{-6} \text{ m s}^{-1}$; this is unprecedented accuracy for tracking data at solar conjunction.

value, and therefore unappreciable) is, however, accounted for by the model. The variability of the solar constant, and possible small changes in the thermo-optical coefficients due to ageing and temperature variations, are insignificant.

The question at what level can violations of general relativity be expected, does not have a satisfactory answer yet. A long-range scalar field is currently assumed to have a fundamental role in primordial cosmology: although it decays with the expansion of the Universe, its present remnant would entail not only violations of the two main tests of general relativity, but also a lack of universality of the constants of microphysics, as assumed in the equivalence principle^{7,8}. A claim, based on quasar absorption lines, that the fine-structure constant α was weaker in the distant past has been made recently, thereby violating the equivalence principle^{16–19}. If this claim is confirmed, and reconciled with other constraints on the variation of fundamental constants²⁰, such a finding would be the first serious challenge to Einstein's model. No detailed theory is available about the expected amounts of these violations, but $\gamma - 1$ should be negative and, possibly, in the range 10^{-5} – 10^{-7} . Therefore, our result with accuracy not far from this range places an important constraint on this cosmological scenario. $\square$

Method

The dynamical model used in the orbital fit is particularly simple, thanks to the large distance from the Sun, the location of the spacecraft in interplanetary space and the lack of unknown gravitational perturbations by Solar System bodies. The Jet Propulsion Laboratory's Orbit Determination Program has been used in the integration of the equations of motion and the orbital solution, based on planetary ephemerides and ancillary data, such as station location and Earth orientation parameters. To speed up the data processing, the observables have been compressed at $\tau = 300$ s by differencing the detected phases. Owing to the spectral characteristics of the noise (Supplementary Fig. S2), the data compression does not in any way affect the final result. We have used up to 12 free 'solve-for' parameters: (1) the six components of the state vector at the start of the experiment; (2) the three components of the non-gravitational acceleration due to the RTGs in the spacecraft frame; (3) the specular and diffuse reflectivity of the high-gain antenna, which determine the magnitude (and the direction) of the non-gravitational acceleration owing to solar radiation pressure; (4) the relativistic parameter γ . 'Consider' parameters (quantities not solved-for, but whose uncertainty is taken into account in the solution) include the dry troposphere, the station location, polar motion and the Earth Love numbers (which intervene in the solid tide model).

As discussed above, among the five parameters that control the non-gravitational acceleration, three (the non-radial components of the thermal thrust from the RTGs and one of the two optical coefficients of the high-gain antenna) are poorly determined. It is therefore appropriate to investigate a solution including only the other two, namely the radial acceleration due to the RTGs and the diffuse reflectivity of the antenna; by so doing, most non-gravitational perturbations are accounted for at a level consistent with the accuracy of the tracking data. The value of the other three parameters, the specular reflectivity and the non-radial components of the RTGs acceleration, together with their uncertainties have been taken from a separate fit carried out on the data from the Cassini solar opposition experiment, as mentioned above. This is our main orbital fit, with only nine parameters to be determined. The a priori uncertainty of the parameters that are not estimated, but affect the solution (such as the station geocentric coordinates and those derived from the gravitational wave experiment), has been included in the computation of the covariance matrix.

The data are assumed to be independent, but for each passage they are weighted with their own standard deviation. The variability of the results has been explored with different assumptions, in particular: a change in the threshold for discarding the outliers; different sampling time; and fitting separately the data before and after the conjunction. These trials clarified several issues including the structure of the covariance matrix, but did not in any way mar the final result.

The residuals of the orbital fit (Fig. 3) show a remarkably white spectrum (see Supplementary Fig. S2), which corresponds to the $\sim 1/\sqrt{\tau}$ observed dependence of both the root-mean-square (r.m.s.) deviation and the Allan deviation from the sampling time τ used in the orbital fit. After obvious outliers are removed (mostly introduced by incorrect tropospheric calibrations), the statistical distribution of the data shows a clearly gaussian behaviour.

We have also explored the full 12-parameter orbital fit and obtained similar results, with $\gamma = 1 + (1.35 \pm 2.47) \times 10^{-5}$, and no appreciable variation of the r.m.s. value of the residuals. Reassuringly, the non-gravitational accelerations so obtained are also fully consistent with the value obtained from the previous opposition experiment and with a long arc orbital solution generated by the Cassini Navigation Team using a data set spanning almost two years. In another trial, in addition to the spacecraft state vector we used just two free parameters— γ and a single, radial non-gravitational acceleration—and obtained the result $\gamma = 1 + (0.21 \pm 2.43) \times 10^{-5}$, with a small increase in the r.m.s. value of the residuals. The understanding of the physics involved and the number of different checks that have been performed have increased our confidence in the results.

Received 25 February; accepted 15 August 2003; doi:10.1038/nature01997.

- Will, C. M. *Theory and Experiment in Gravitational Physics* (Cambridge Univ. Press, Cambridge, 1993).
- Will, C. M. The confrontation between General Relativity and experiment. *Living Rev. Relativity* **4**, 4 (cited on 15 Aug 2003); at (<http://www.livingreviews.org/lrr-2001-4>).
- Ciufolini, I. & Wheeler, J. A. *Gravitation and Inertia* (Princeton Univ. Press, Princeton, 1995).
- Dyson, F. W., Eddington, A. S. & Davidson, C. A determination of the deflection of light by the Sun's gravitational field, from observations made at the total eclipse of May 29, 1919. *Phil. Trans. R. Soc. Lond. A* **220**, 291–333 (1920).
- Lebach, D. E. *et al.* Measurement of the solar gravitational deflection of radio waves using very-long-baseline interferometry. *Phys. Rev. Lett.* **75**, 1439–1442 (1995).
- Reasenber, R. D. *et al.* Viking relativity experiment: verification of signal retardation by solar gravity. *Astrophys. J.* **234**, L219–L221 (1979).
- Damour, T. & Polyakov, A. M. The string dilaton and a least coupling principle. *Nucl. Phys. B* **423**, 532–558 (1994).
- Damour, T., Piazza, F. & Veneziano, G. Violations of the equivalence principle in a dilaton-runaway scenario. *Phys. Rev. D* **66**, 046007 (2002).
- Iess, L., Giampieri, G., Anderson, J. D. & Bertotti, B. Doppler measurement of the solar gravitational deflection. *Class. Quant. Grav.* **16**, 1487–1502 (1999).
- Bertotti, B. & Giampieri, G. Relativistic effects for Doppler measurements near solar conjunction. *Class. Quant. Grav.* **9**, 777–793 (1992).
- Bertotti, B., Comoretto, G. & Iess, L. Doppler tracking of spacecraft with multifrequency links. *Astron. Astrophys.* **269**, 608–616 (1993).
- Bertotti, B. & Giampieri, G. Solar coronal plasma in Doppler measurements. *Solar Phys.* **178**, 523–545 (1998).
- Iess, L. *et al.* in *Aerospace Conf. Proc. (Montana, March 2003)* Vol. 1, 205–211 (Institute of Electrical and Electronics Engineers (IEEE), 2003).
- Tortora, P., Iess, L. & Ekelund, J. E. in *Proc. World Space Congr. (Houston, 2002)* Paper IAC-02-Q-2.03 (AIAA, Reston, VA, 2002).
- Tortora, P., Iess, L. & Herrera, R. G. in *Aerospace Conf. Proc. (Montana, March 2003)* Vol. 3, 1465–1473 (Institute of Electrical and Electronics Engineers (IEEE), 2003).
- Murphy, M. T. *et al.* Possible evidence for a variable fine-structure constant from QSO absorption lines: motivations, analysis and results. *Mon. Not. R. Astron. Soc.* **327**, 1208–1222 (2001).
- Murphy, M. T. *et al.* Possible evidence for a variable fine-structure constant from QSO absorption lines: systematic errors. *Mon. Not. R. Astron. Soc.* **327**, 1223–1236 (2001).
- Murphy, M. T. *et al.* Further constraints on variation of the fine-structure constant from alkali-doublet QSO absorption lines. *Mon. Not. R. Astron. Soc.* **327**, 1237–1243 (2001).
- Murphy, M. T. *et al.* Improved constraints on possible variations of physical constants from HI 21-cm and molecular QSO absorption lines. *Mon. Not. R. Astron. Soc.* **327**, 1244–1248 (2001).
- Damour, T. String theory, cosmology and varying constants. *Astrophys. Space Sci.* **283**, 445–456 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The work of B.B., L.I. and P.T. has been funded by the Italian Space Agency (ASI). The Jet Propulsion Laboratory Radio Science Systems Group, the engineers and staff of the DSS25 station, and the Cassini Project have been crucial for the success of this experiment. We thank J. W. Armstrong, J. J. Bordin and D. C. Roth for their contributions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.I. (iess@hermes.ing.uniroma1.it).

Uniform resonant chaotic mixing in fluid flows

T. H. Solomon¹ & Igor Mezić²

¹Department of Physics, Bucknell University, Lewisburg, Pennsylvania 17837, USA

²Department of Mechanical and Environmental Engineering, University of California at Santa Barbara, Santa Barbara, California 93106, USA

Laminar flows can produce particle trajectories that are chaotic^{1,2}, with nearby tracers separating exponentially in time. For time-periodic, two-dimensional flows and steady three-dimensional (3D) flows, enhancements in mixing due to chaotic advection are typically limited by impenetrable transport barriers that form at the boundaries between ordered and chaotic mixing regions. However, for time-dependent 3D flows, it has been proposed theoretically^{3–5} that completely uniform mixing is possible through a resonant mechanism⁵ called singularity-

induced diffusion; this is thought to be the case even if the time-dependent and 3D perturbations are infinitesimally small. It is important to establish the conditions for which uniform mixing is possible and whether or not those conditions are met in flows that typically occur in nature. Here we report experimental and numerical studies of mixing in a laminar vortex flow that is weakly 3D and weakly time-periodic. The system is an oscillating horizontal vortex chain (produced by a magnetohydrodynamic technique) with a weak vertical secondary flow that is forced spontaneously by Ekman pumping—a mechanism common in vortical flows with rigid boundaries, occurring in many geophysical, industrial and biophysical flows. We observe completely uniform mixing, as predicted^{3–5} by singularity-induced diffusion, but only for oscillation periods close to typical circulation times.

Most previous studies of chaotic advection have focused on two-dimensional (2D), time-periodic flows. Transport barriers are prevalent in those flows⁶; in fact, chaotic transport barriers have been proposed as mechanisms for chemical isolation in geophysical flows^{7,8}, such as the Antarctic circumpolar region (the ozone hole) and Jupiter's Great Red Spot. Recent studies have further proposed that chaotic advection may have a pivotal role in the expanding field of microfluidic devices⁹ and may also be important for many

biophysical processes^{10,11}. The possibility of chaotic motion of tracers in 3D flows was first suggested in 1966 (ref. 12) and was pursued numerically by another study¹³. However, systematic investigations of chaotic mixing in laminar 3D flows have only begun to appear in the literature during the past decade^{3,4,14–24}. Further studies^{3,4} extended the original work to show that uniform, barrier-free transport is possible through a mechanism called 'singularity-induced diffusion' (SID), even if the 3D and time-dependent perturbations are both infinitesimally small. A recent study⁵ showed that a resonance mechanism is responsible for SID. When typical circulation frequencies are resonant with the driving frequency, saddle foci-type orbits enable tracers to spiral off one adiabatic surface and onto a different one, producing global mixing. An implication of this resonance picture, however, is that mixing will not be uniform if there are regions in the flow where the resonance condition is not met.

(Note that as SID—which applies to volume-preserving, time-dependent 3D flows—is enabled by spiral-node fixed points, it differs from another mechanism for global transport called 'Arnol'd diffusion', occurring in hamiltonian systems with more than 2 degrees of freedom.)

The flow studied here (Fig. 1) is dominated by a horizontal chain of alternating vortices with a secondary flow due to Ekman pump-

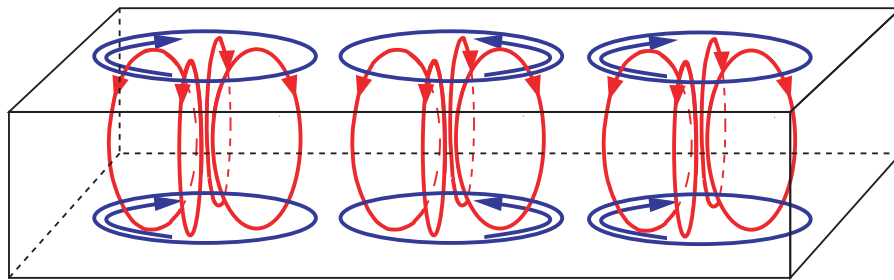


Figure 1 Sketch of a portion of the flow. The primary flow (blue) is a chain of alternating horizontal vortices; a weak secondary flow (red) is generated naturally by Ekman pumping, which draws fluid inward along the bottom of the vortices and up the vortex centres.

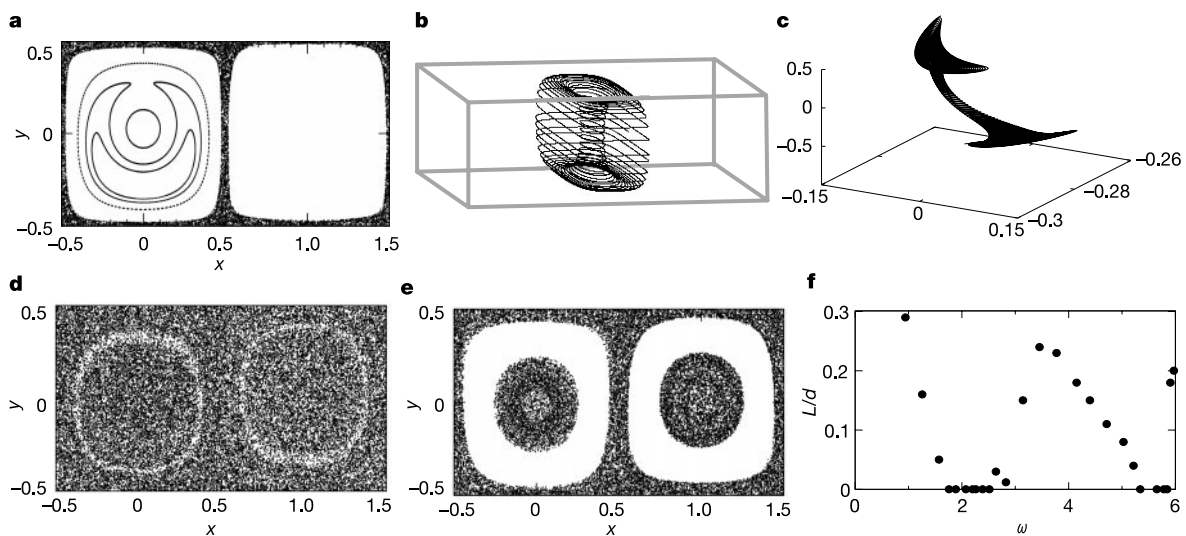


Figure 2 Numerical simulations of equations (1)–(3). **a**, Poincaré section for 2D, time-periodic case ($\epsilon = 0$, $b = 0.01$, $\omega = 2.5$), determined by plotting locations of five tracers once each period, initially located on the x axis at $x = 0.49, 0.4, 0.3, 0.2$ and 0.1 . **b**, Trajectory of tracer in 3D, time-independent flow ($\epsilon = 0.005$, $b = 0$). **c**, Poincaré section for a tracer near a periodic orbit for 3D, time-periodic case ($\epsilon = 0.0001$, $b = 0.001$, $\omega = 2.5$), showing the tracer spiralling off one adiabatic surface (the spiral is very tightly wound and may be difficult to resolve) and onto a different one. **d**, Horizontal

slice ($-0.1 < z < 0.1$) of Poincaré section of a single tracer for 3D, time-periodic case ($\epsilon = 0.005$, $b = 0.01$, $\omega = 2.5$); the tracer is initially located at $(x, y, z) = (0.49, 0, 0)$, and the forcing frequency is almost resonant with tracers circulating in the central torus. **e**, Horizontal slice of Poincaré section for a non-resonant forcing frequency ($\epsilon = 0.005$, $b = 0.01$, $\omega = 4.0$). **f**, Width L of the excluded region (as a fraction of the vortex width d) versus non-dimensional driving frequency ($\epsilon = 0.005$, $b = 0.01$). Uniform mixing is achieved for the frequencies where the width goes to zero.

ing, a process that occurs whenever a vortical flow is bounded by a solid surface. Radial pressure gradients due to the no-slip boundary condition push the fluid inward just above the solid boundary and up through the vortex centres. This is a common 3D flow perturbation; we therefore expect the internal mixing properties observed with this flow to be generic to a wide variety of vortical flows: basically, laminar vortex flows in the presence of a rigid boundary.

Time dependence takes the form of lateral oscillations of the vortex chain, similar to the oscillatory instability of Rayleigh–Bénard convection²⁵. We have developed a model that captures the essential features of this flow:

$$v_x = dx/dt = -\cos(\pi x_s(t))\sin(\pi y) + \epsilon \sin(2\pi x_s(t))\sin(\pi z) \quad (1)$$

$$v_y = dy/dt = \sin(\pi x_s(t))\cos(\pi y) + \epsilon \sin(2\pi y)\sin(\pi z) \quad (2)$$

$$v_z = dz/dt = 2\epsilon \cos(\pi z)[\cos(2\pi x_s(t)) + \cos(2\pi y)] \quad (3)$$

Distances are scaled by the vortex width d , and time is scaled by the advective time d/U , where U is the maximum flow velocity. The lateral oscillation is accounted for by a shifted x -coordinate $x_s(t) = x + b\sin\omega t$, where b and ω are the non-dimensionalized oscillation amplitude and frequency. The strength of the secondary (3D) component of the flow is characterized by ϵ . This model is not intended to be a rigorous description of the flow in our experiments; rather, it is the simplest phenomenological model that captures the essential features of an alternating vortex flow with weak Ekman pumping. In particular, the model equations assume free-slip boundary conditions—a model with no-slip boundary conditions would have significantly more complicated y - and z -dependence^{26,27}. Nevertheless, this simplified model successfully captures the dominant features of mixing in the experiments, as is shown below.

Numerical trajectories for the model are determined by integrat-

ing equations (1)–(3) using a fourth-order Runge–Kutta technique. The results are shown in Fig. 2. The 2D case ($\epsilon = 0$) is shown in Fig. 2a; trajectories are ordered in most of the flow, except for a chaotic band around and between the vortices. A tracer within the chaotic region never crosses into the ordered region in the vortex middle. A tracer trajectory for a 3D, time-independent flow is shown in Fig. 2b. An impurity near a vortex edge spirals upward, inward and then down through the centre and back out again. During this motion, the tracer visits different slices of the base 2D flow at different radii with different circulation times.

The mechanism for SID discussed in ref. 5 is seen numerically in Fig. 2c for a flow with weak time dependence and weak three-dimensionality. Near a resonance, a tracer spirals off what would have been an adiabatic surface (in the absence of SID) and onto a different surface. The result is nearly uniform mixing, as seen in Fig. 2d. But this mechanism works only if the frequency of oscillation is such that every tracer eventually reaches a radius with circulation frequency resonant with this driving frequency. Practically, this means that the oscillation frequency must be resonant with the circulation frequency for tracers in the lightened region in Fig. 2d. If forced at a different frequency, a toroidal region forms into and out of which there is zero or little mixing (Fig. 2e).

This resonant behaviour is plotted versus frequency in Fig. 2f; the frequencies where the width of the excluded region vanishes are those where uniform mixing should be expected. Practically, the frequency range for uniform mixing is slightly wider than shown in Fig. 2f, as molecular diffusion can mix impurities into the excluded region if the width is small. These results indicate a signature that should be seen in the experiments; namely, transport should be uniform or nearly uniform if forced at a frequency within or near these resonant bands, but should be characterized by toroidal regions of weak mixing if forced at different frequencies.

Experimentally, the flow in Fig. 1 is generated by a magneto-

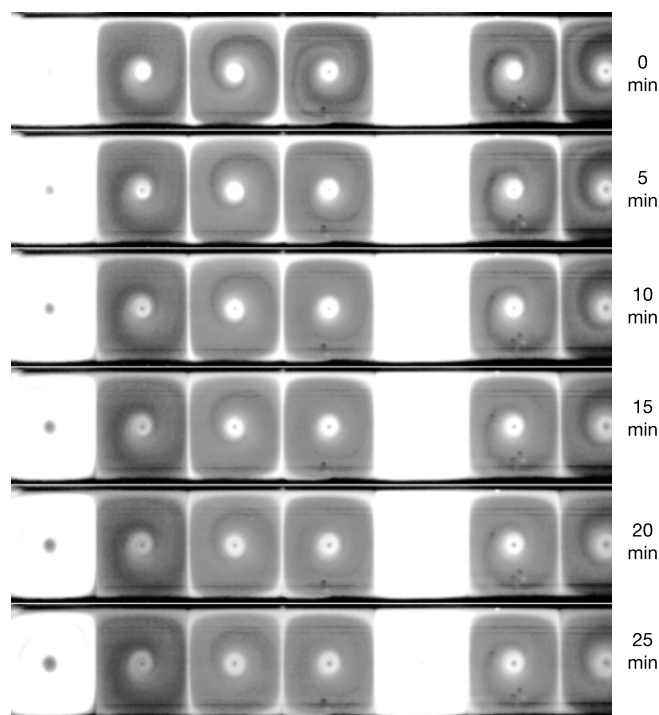


Figure 3 Experimental images of mixing for time-independent flow ($b = 0$). The same enhancement is used for all the images in this sequence. The images are saturated at large concentration values; consequently, decreases in the concentration of the injection vortices may not be noticeable in this figure or in Figs 4 and 5. For the time-independent case shown here, mixing is very slow both within and between the vortices. Concentration profiles for the centre vortex are available as Supplementary Information.

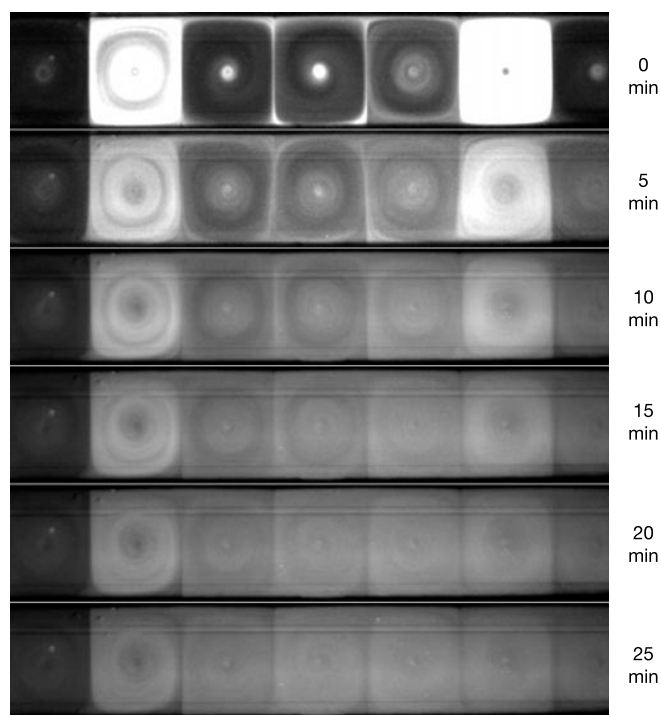


Figure 4 Experimental images of mixing for time-periodic flow forced near resonance ($b = 0.02$, 0.10 Hz oscillation frequency, corresponding to a non-dimensional angular frequency $\omega = 2.5$). The concentration becomes homogenized within the vortices. Concentration profiles for the centre vortex are available as Supplementary Information.

hydrodynamic technique^{28,29} in which an electric current passes through a thin (2 mm) layer of dilute (0.006 M) H_2SO_4 with a free surface. The region of interest has horizontal dimensions 2.2×22 cm. The current interacts with an alternating magnetic field produced by magnets below the fluid; the result is a chain of vortices with a maximum velocity of approximately 7 mm s^{-1} . Ekman pumping is generated spontaneously, owing to a glass plate below the vortices. For the impurity, we use fluorescent latex microspheres (0.103- μm diameter) with a diffusion coefficient $4.5 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$, based on the Stokes–Einstein relation³⁰, small enough to minimize diffusive mixing. The flow is illuminated with black light, and the fluorescing microspheres are imaged by a CCD video camera.

When viewed from the side, the impurity can be seen moving inward slowly along the bottom and up through the centres of the vortices; from observations, it takes approximately ten horizontal circulation times (as measured for tracers circling in the outer portions of the vortices) for the impurity to circulate appreciably in the vertical direction owing to Ekman pumping. Images of mixing from above for a time-independent, 3D flow are shown in Fig. 3. In each experiment, we inject the impurity into two different vortices. During the injection process, some of the impurity slips out of these two vortices and into the edges of the neighbouring vortices. Left undisturbed (and with no time dependence) for 20 min, Ekman pumping carries some of the impurity into the centres of the vortices. The result is the ‘ $t = 0$ min’ case in which most vortices have a thin line of impurity around the edges and spots in the centre. For Fig. 3, we leave the system undisturbed after this initial condition. The result for this time-independent case is very slow internal mixing.

Figure 4 shows mixing for the case with weak time-periodic forcing with a frequency resonant with the typical circulation frequencies. Within about 25 min (approximately 150 oscillation periods), the impurity concentration is almost completely uniform within the vortices. The mixing is different in a subtle but important

way if the frequency of the time dependence is in a non-resonant band (Fig. 5). There is still significant mixing within the vortices. However, there are toroidal regions into which impurity does not mix readily: these regions appear in Fig. 5 as darkened annuli, and they persist for over an hour. These darkened annuli are the excluded toroidal regions seen in the simulations (compare Fig. 5 with Fig. 2e).

For the run shown in Fig. 5, mixing of the impurity out of the two injection vortices is also retarded by the transport barriers. For later times (not shown), higher concentrations persist in the tori in these two vortices; in fact, a brightened torus is still clear after 2 h in the left-most vortex. At this point, we change the oscillation frequency back to the original resonant value, and the brightened region quickly dissipates (in 5–10 min), further evidence of the resonant nature of this internal mixing. (See Supplementary Information for a video of this, along with videos of the sequences of Figs 4 and 5.)

We find that mixing in a weakly 3D, weakly time-periodic flow is nearly uniform, but only if forced at a period resonant with the typical circulation times; specifically, circulation times at radii corresponding to the centres of what would otherwise be the excluded tori. This restriction might seem to limit the applicability of SID as a mechanism for uniform mixing in real flows; however, in nature, flows often choose precisely this circulation frequency for time-periodic instabilities. In fact, in the experiments presented here, it is experimentally challenging to avoid internal mixing because even very slight time-dependent perturbations due to aggregation of the latex microspheres or due to any advected floating scum in the system results in significantly enhanced internal mixing. Both of these mechanisms result in time dependence with frequencies determined by circulation frequencies, so the resonant condition is satisfied automatically. As time-dependent perturbations are often linked to circulation, we expect uniform mixing to be common in many natural vortex flows with Ekman pumping. □

Received 30 June; accepted 11 August 2003; doi:10.1038/nature01993.

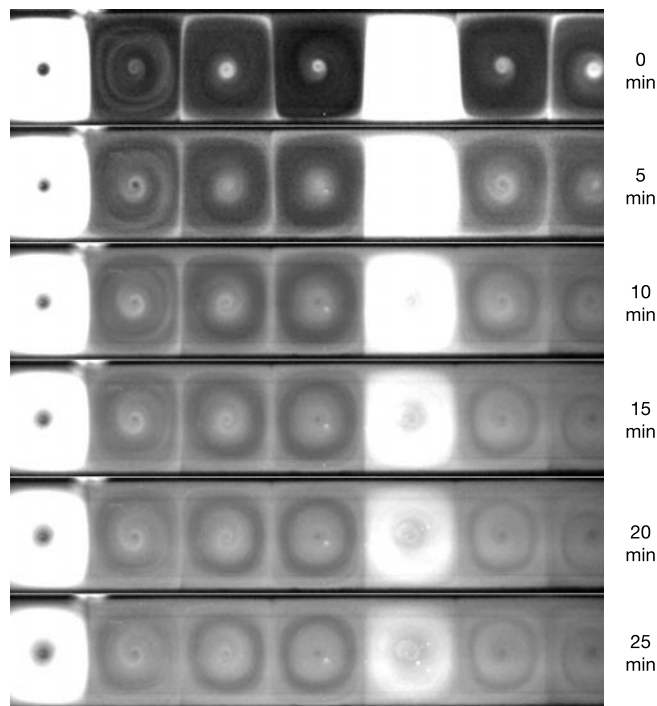


Figure 5 Experimental images of mixing for time-periodic flow forced away from a resonant band ($b = 0.02$, 0.160 Hz oscillation frequency, corresponding to a non-dimensional angular frequency $\omega = 4.2$). Darkened annular bands are seen in the later images, indicating regions into which mixing is weak. Concentration profiles for the centre vortex are available as Supplementary Information.

1. Aref, H. Stirring by chaotic advection. *J. Fluid Mech.* **143**, 1–21 (1984).
2. Aref, H. Fluid dynamics: Order in chaos. *Nature* **401**, 756–758 (1999).
3. Feingold, M., Kadanoff, L. P. & Piro, O. Passive scalars, 3-dimensional volume-preserving maps and chaos. *J. Stat. Phys.* **50**, 529–565 (1988).
4. Cartwright, J. H. E., Feingold, M. & Piro, O. Chaotic advection in three-dimensional unsteady incompressible laminar flow. *J. Fluid Mech.* **316**, 259–284 (1996).
5. Mezić, I. Break-up of invariant surfaces in action-angle maps and flows. *Physica D* **154**, 51–67 (2001).
6. Rom-Kedar, V., Leonard, A. & Wiggins, S. An analytical study of transport, mixing and chaos in an unsteady vortical flow. *J. Fluid Mech.* **214**, 347–394 (1990).
7. Behringer, R. P., Meyers, S. D. & Swinney, H. L. Chaos and mixing in a geostrophic flow. *Phys. Fluids A* **3**, 1243–1249 (1991).
8. Pierrehumbert, R. T. Large scale horizontal mixing in planetary atmospheres. *Phys. Fluids A* **3**, 1250–1260 (1991).
9. Stroock, A. D., Dertinger, S. K. W., Mezić, I., Stone, H. A. & Whitesides, G. M. Chaotic mixer for microchannels. *Science* **295**, 647–651 (2002).
10. Karolyi, G., Pentek, A., Scheuring, I., Tel, T. & Toroczkai, Z. Chaotic flow: The physics of species coexistence. *Proc. Natl Acad. Sci. USA* **25**, 13661–13665 (2000).
11. Swanson, P. D., Muzzio, F. J., Annappagada, A. & Adjei, A. Numerical analysis of motion and deposition of particles in cascade impactors. *Int. J. Pharm.* **142**, 33–51 (1996).
12. Arnold, V. I. Sur la géométrie différentielle des groupes de lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. *Ann. Inst. Fourier* **16**, 316–361 (1966).
13. Hénon, M. Sur la topologie des lignes de courant dans un cas particulier. *C. R. Acad. Sci. Paris A* **262**, 312–314 (1966).
14. Bajer, K. & Moffatt, H. K. On a class of steady confined Stokes flows with chaotic streamlines. *J. Fluid Mech.* **212**, 337–363 (1990).
15. Stone, H. A., Nadim, A. & Strogatz, S. H. Chaotic streaklines inside drops immersed in steady linear flows. *J. Fluid Mech.* **232**, 629–646 (1991).
16. Holm, D. D. & Kimura, Y. Zero-helicity Lagrangian kinematics of three dimensional advection. *Phys. Fluids A* **3**, 1033–1038 (1991).
17. Mackay, R. S. Transport in 3-D volume-preserving flows. *J. Nonlinear Sci.* **4**, 329–354 (1994).
18. Mezić, I. & Wiggins, S. On the integrability and perturbation of three dimensional fluid flows with symmetry. *J. Nonlinear Sci.* **4**, 157–194 (1994).
19. Ashwin, P. & King, G. P. Streamline topology in eccentric Taylor vortex flow. *J. Fluid Mech.* **285**, 215–247 (1995).
20. Fountain, G. O., Khakhar, D. V., Mezić, I. & Ottino, J. M. Chaotic mixing in a bounded 3D flow. *J. Fluid Mech.* **417**, 265–301 (2000).
21. Sotiropoulos, F., Ventikos, Y. & Lackey, T. C. Chaotic advection in 3-D stationary vortex-breakdown bubbles: Šil'nikov's chaos and the devil's staircase. *J. Fluid Mech.* **444**, 257–297 (2001).

22. Fountain, G. O., Khakhar, D. V. & Ottino, J. M. Visualization of three-dimensional chaos. *Science* **281**, 683–686 (1998).
23. Shinbrot, T., Alvarez, M. M., Zalc, J. M. & Muzzio, F. J. Attraction of minute particles to invariant regions of volume preserving flows by transients. *Phys. Rev. Lett.* **86**, 1207–1210 (2001).
24. Vainshtein, D. L., Vasiliev, A. A. & Neishtadt, A. J. Changes in the adiabatic invariant and streamline chaos in confined incompressible Stokes flow. *Chaos* **6**, 67–77 (1996).
25. Clever, R. M. & Busse, F. H. Transition to time-dependent convection. *J. Fluid Mech.* **65**, 625–645 (1974).
26. Solomon, T. H. & Gollub, J. P. Chaotic particle transport in time-dependent Rayleigh-Bénard convection. *Phys. Rev. A* **38**, 6280–6286 (1988).
27. Chandrasekhar, S. *Hydrodynamic and Hydromagnetic Stability* 39 (Clarendon, Oxford, 1961).
28. Solomon, T. H., Tomas, S. & Warner, J. L. The role of lobes in chaotic mixing of miscible and immiscible impurities. *Phys. Rev. Lett.* **77**, 2682–2685 (1996).
29. Solomon, T. H., Tomas, S. & Warner, J. L. Chaotic mixing of immiscible and immiscible impurities in a two-dimensional flow. *Phys. Fluids* **10**, 342–350 (1998).
30. Cussler, E. L. *Diffusion: Mass Transfer in Fluid Systems* (Cambridge Univ. Press, Cambridge, 1984).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by the US National Science Foundation, AFOSR, and by a Sloan Fellowship grant.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.S. (tsolomon@bucknell.edu) or I.M. (mezic@engineering.ucsb.edu).

Microwave oscillations of a nanomagnet driven by a spin-polarized current

S. I. Kiselev^{1*}, J. C. Sankey^{1*}, I. N. Krivorotov¹, N. C. Emley¹, R. J. Schoelkopf², R. A. Buhrman¹ & D. C. Ralph¹

¹Cornell University, Ithaca, New York 14853, USA

²Department of Applied Physics and Physics, Yale University, New Haven, Connecticut 06511, USA

* These authors contributed equally to this work

The recent discovery that a spin-polarized electrical current can apply a large torque to a ferromagnet, through direct transfer of spin angular momentum, offers the possibility of manipulating magnetic-device elements without applying cumbersome magnetic fields^{1–16}. However, a central question remains unresolved: what type of magnetic motions can be generated by this torque? Theory predicts that spin transfer may be able to drive a nanomagnet into types of oscillatory magnetic modes not attainable with magnetic fields alone^{1–3}, but existing measurement techniques have provided only indirect evidence for dynamical states^{4,6–8,12,14–16}. The nature of the possible motions has not been determined. Here we demonstrate a technique that allows direct electrical measurements of microwave-frequency dynamics in individual nanomagnets, propelled by a d.c. spin-polarized current. We show that spin transfer can produce several different types of magnetic excitation. Although there is no mechanical motion, a simple magnetic-multilayer structure acts like a nanoscale motor; it converts energy from a d.c. electrical current into high-frequency magnetic rotations that might be applied in new devices including microwave sources and resonators.

We examine samples made by sputtering a multilayer of composition 80 nm Cu/40 nm Co/10 nm Cu/3 nm Co/2 nm Cu/30 nm Pt onto an oxidized silicon wafer and then milling through part of the multilayer (Fig. 1a) to form a pillar with an elliptical cross-section of lithographic dimensions 130 nm × 70 nm (ref. 17). Top

contact is made with a Cu electrode. Transmission or reflection of electrons from the thicker ‘fixed’ Co layer produces a spin-polarized current that can apply a torque to the thinner ‘free’ Co layer. Subsequent oscillations of the free-layer magnetization relative to the fixed layer change the device resistance¹⁸ so, under conditions of d.c. current bias, magnetic dynamics produce a time-varying voltage (with typical frequencies in the microwave range). If the oscillations were exactly symmetric relative to the direction of the fixed-layer moment, voltage signals would occur only at multiples of twice the fundamental oscillation frequency, f . To produce signal strength at f , we apply static magnetic fields (H) in the sample plane a few degrees away from the magnetically easy axis of the free layer. All data are taken at room temperature, and by convention positive current I denotes electron flow from the free to the fixed layer.

In characterization measurements done at frequencies <1 kHz, the samples exhibit the same spin-transfer-driven changes in resistance reported in previous experiments^{7,9} (Fig. 1b). For H smaller than the coercive field of the free layer ($H_c \approx 600$ Oe), an applied current produces hysteretic switching of the magnetic layers between the low-resistance parallel (P) and high-resistance anti-parallel (AP) states. Sweeping H can also drive switching between the P and AP states (Fig. 1b, inset). For H larger than 600 Oe, the current produces peaks in the differential resistance dV/dI that have been assumed previously to be associated with dynamical magnetic excitations^{4,6–8}. The resistance values displayed in Fig. 1b include a lead resistance of $\sim 6 \Omega$ from high-frequency (50 GHz) probes and a top-contact resistance of $\sim 9 \Omega$.

We measure the spectra of microwave power that result from magnetic motions by using a heterodyne mixer circuit¹⁹ (Fig. 1a). This circuit differs from the only previous experiment to probe spin-transfer-driven magnetic oscillations⁸ in that the sample is not exposed to a large high-frequency magnetic field that would alter its dynamics. The filter on the output of our mixer passes 25–100 MHz, giving a frequency resolution of ~ 200 MHz. We calibrate the circuit by measuring temperature-dependent Johnson noise from test resistors. When we state values of emitted power, they will correspond to the power available to a load matched to the sample resistance, R . To convert to the power delivered to a 50- Ω line, one should multiply our values by the power transmission coefficient $1 - \Gamma^2 = 1 - [(R - 50 \Omega)/(R + 50 \Omega)]^2$.

We first consider the microwave spectrum from sample 1 for $H = 2$ kOe. For both negative I and small positive I we measure only frequency-independent Johnson noise. We will subtract this background from all the spectra we display. At $I = 2.0$ mA, we begin to resolve a microwave signal at 16.0 GHz (Fig. 1c, d). A second-harmonic peak is also present (Fig. 1c, inset). As I is increased, these initial signals grow until $I \approx 2.4$ mA, beyond which the dynamics change to a different regime (Fig. 1d). In Fig. 1e, we compare the H -dependence of the measured frequency for the initial signals to the formula for small-angle elliptical precession of a thin-film ferromagnet²⁰:

$$f = \frac{\gamma}{2\pi} \sqrt{(H + H_{\text{an}} + H_{\text{d}})(H + H_{\text{an}} + H_{\text{d}} + 4\pi M_{\text{eff}})} \quad (1)$$

Here γ is the gyromagnetic ratio, H_{an} accounts for a uniaxial easy-axis anisotropy, H_{d} models the coupling from the fixed layer, and $4\pi M_{\text{eff}} = 4\pi M_{\text{s}} - 2K_{\text{u}}/M_{\text{s}}$, with M_{s} the saturation magnetization and K_{u} a uniaxial perpendicular anisotropy²¹. The fit is excellent and gives the values $4\pi M_{\text{eff}} = 6.8 \pm 0.1$ kOe and $H_{\text{an}} + H_{\text{d}} = 1.18 \pm 0.04$ kOe. The value for $4\pi M_{\text{eff}}$ is less than $4\pi M_{\text{s}}$ for bulk Co (16 kOe) as expected due to significant perpendicular anisotropy in Co/Cu(111) films (see Fig. 3 in ref. 22). Similar fits for other samples yield $4\pi M_{\text{eff}}$ in the range 6.7–12 kOe. Superconducting quantum interference device (SQUID) measurements on test samples containing many 3-nm Co layers give $4\pi M_{\text{eff}} = 10 \pm 1$ kOe.

On the basis of the agreement with equation (1) we identify the initial signals as arising from small-angle elliptical precession of the free layer, thereby confirming pioneering predictions that spin-transfer can coherently excite this uniform spin-wave mode². We can make a rough estimate for the amplitude of the precession angle, $\theta_{\max}$, and the misalignment θ_{mis} between the precession axis and the fixed-layer moment (induced by the applied field), based on the integrated microwave power measured about f and $2f$ (P_f and P_{2f}). Assuming for simplicity that $\theta(t) = \theta_{\text{mis}} + \theta_{\max}\sin(\omega t)$, that

the angular variation in resistance $\Delta R(\theta) = \Delta R_{\max}(1 - \cos(\theta))/2$, and that $|\theta_{\text{mis}} \pm \theta_{\max}| \ll 1$, we calculate:

$$\theta_{\max}^4 \approx \frac{512P_{2f}R}{\Delta R_{\max}^2 I^2} \quad (2)$$

$$\theta_{\text{mis}}^2 \approx \frac{32P_f R}{\Delta R_{\max}^2 I^2 \theta_{\max}^2} \quad (3)$$

where $R = 12.8 \Omega$ and $\Delta R_{\max} = 0.11 \Omega$ is the resistance change between P and AP states. For the spectrum from sample 1 in the inset to Fig. 1c, we estimate that $\theta_{\text{mis}} \approx 9^\circ$, and the precessional signal first becomes measurable above background when $\theta_{\max} \approx 10^\circ$.

With increasing currents, the nanomagnet exhibits additional dynamical regimes. As I is increased beyond 2.4 mA to 3.6 mA for sample 1, the microwave power grows by two orders of magnitude, peak frequencies shift abruptly, and the spectrum acquires a significant low-frequency background (Fig. 1c). In many samples (including sample 2 below) the background becomes so large that some spectral peaks are difficult to distinguish. Within this large-amplitude regime, peaks shift down in frequency with increasing current (Fig. 1f). The large-amplitude signals persist for I up to 6.0 mA, where the microwave power plummets sharply at the same current for which there is a shoulder in dV/dI . The state that appears thereafter has a d.c. resistance 0.04Ω lower than the AP state and

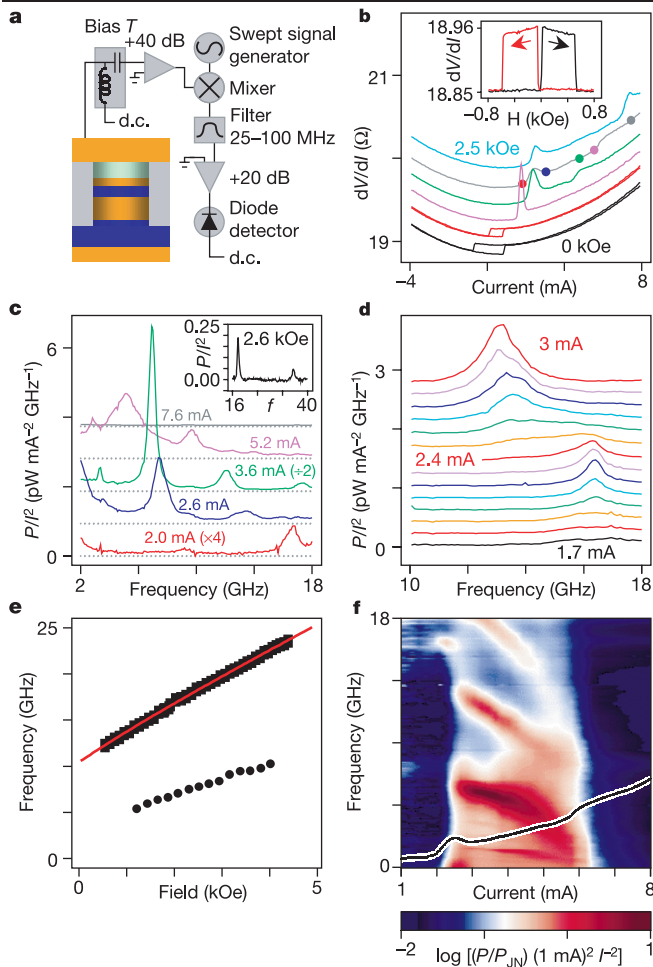


Figure 1 Resistance and microwave data for sample 1. **a**, Schematic of the sample with copper layers (orange), cobalt (blue), platinum (green) and SiO_2 insulator (grey), together with the heterodyne mixer circuit. Different preamplifiers and mixers allow measurements over 0.5–18 GHz or 18–40 GHz. **b**, Differential resistance versus current for magnetic fields of 0 (bottom), 0.5, 1.0, 1.5, 2.0 and 2.5 kOe (top), with current sweeps in both directions. At $H = 0$, the switching currents are $I_c^+ = 0.88$ mA and $I_c^- = -0.71$ mA, and $\Delta R_{\max} = 0.11 \Omega$ between the P and AP states. Coloured dots on the 2 kOe curve correspond to spectra shown in **c**. Inset to **b**, Magnetoresistance near $I = 0$. Red and black indicate different directions of magnetic-field sweep. **c**, Microwave spectra (with Johnson noise subtracted) for $H = 2.0$ kOe, for $I = 2$ mA (bottom), 2.6, 3.6, 5.2 and 7.6 mA (top). We plot power density divided by I^2 to facilitate comparisons of the underlying changes in resistance at different current values. Inset to **c**, Spectrum at $H = 2.6$ kOe and $I = 2.2$ mA, for which both f and $2f$ peaks are visible on the same scan. **d**, Microwave spectra at $H = 2.0$ kOe, for current values from 1.7 to 3.0 mA in 0.1-mA steps, showing the growth of the small-amplitude precessional peak and then a transition in which the second harmonic signal of the large-amplitude regime appears. **e**, Magnetic-field dependence of the small-amplitude signal frequency (top) and the frequency of the fundamental in the large-amplitude regime at $I = 3.6$ mA (bottom). The line is a fit to equation (1). **f**, Microwave power density (in colour scale) versus frequency and current for $H = 2.0$ kOe. The black line shows dV/dI versus I from **b**. P_{JN} is the Johnson-noise power level. The curves in **b**, **c** and **d** are offset vertically.

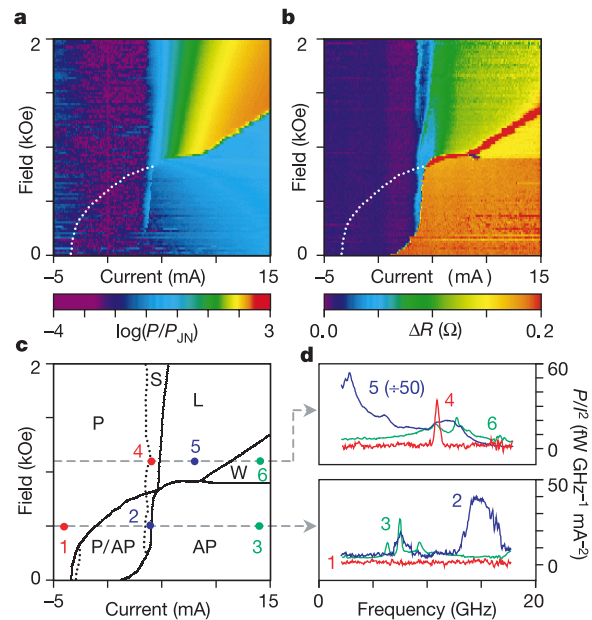


Figure 2 Resistance and microwave data for sample 2. Sample 2 has, at $H = 0$, $I_c^+ = 1.06$ mA, $I_c^- = -3.22$ mA, parallel-state resistance (including top-contact and lead resistances) 17.5Ω , $\Delta R_{\max} = 0.20 \Omega$ between the P and AP states, and $4\pi M_{\text{eff}} = 12$ kOe. **a**, Microwave power above Johnson noise in the frequency range 0.1–18 GHz, plotted in colour scale versus I and H . I is swept from negative to positive values. These data were collected without the mixer circuit by measuring the power with a detector diode after amplification. The dotted white line shows the position of the AP to P transition when I is swept positive to negative. **b**, Differential resistance plotted in colour scale for the same region of I and H . A smooth current-dependent, H -independent background (similar to that of Fig. 1b) is subtracted to better display the different regimes of resistance. Resistance changes are measured relative to the parallel state. **c**, Room-temperature experimental dynamical stability diagram extracted from **a** and **b**. P indicates parallel orientation, AP antiparallel orientation, P/AP parallel/antiparallel bistability, S the small-amplitude precessional regime, L the large-amplitude dynamical regime, and W a state with resistance between P and AP and only small microwave signals. The coloured dots in **c** correspond to the microwave spectra at $H = 500$ and 1,100 Oe shown in **d**.

0.07 Ω above the P state. At even higher current levels (not shown), we sometimes see additional large microwave signals that are not reproducible from sample to sample. These might be associated with dynamics in the fixed layer.

The regions of I and H associated with each type of dynamical mode can be determined by analysing the microwave power and dV/dI (Fig. 2a, b for sample 2). In all eight samples that we have examined in detail, large microwave signals occur for a similarly shaped range of I and H . Samples 1 and 2 exhibit clear structure in dV/dI at the boundaries of the large-amplitude regime, but other samples sometimes lack prominent dV/dI features over part of this border. In Fig. 2c we construct a dynamical stability diagram showing the different modes that can be driven by a d.c. spin-transfer current and a constant in-plane magnetic field. Explaining the existence of all these modes and the positions of their boundaries will provide a rigorous testing ground for theories of spin-transfer-driven magnetic dynamics.

As indicated in Fig. 2c, d, microwave signals can sometimes be observed not only at large H where dynamical modes have been postulated previously^{4,6–8,12,14–16}, but also in the small- H regime of current-driven hysteretic switching. While sweeping to increasing currents at $H = 500$ Oe; for example, microwave peaks corresponding to small-angle precession exist for I within ~ 0.7 mA

below the current for P to AP switching. Similar features are also observed before switching from AP to P at negative bias. We suggest that these microwave signals are due to fluctuations of the free-layer moment away from its easy axis to angles large enough to produce measurable precession, but too small to achieve full reversal over the activation barrier for switching²³. Related signals have been observed recently in magnetic tunnel junctions²⁴.

To understand what type of motion may be associated with the different dynamical modes, we have computed solutions of the Landau–Lifshitz–Gilbert equation of motion for a single-domain magnet^{25–28}. We employ the form of the spin-transfer torque derived in ref. 1. The calculated zero-temperature dynamical phase diagram is presented in Fig. 3a. We have not attempted to adjust parameters to fit our data, but nevertheless the existence and relative positions of the P, AP and small-angle-precession regimes agree well. The model suggests that the large-amplitude microwave signals correspond to large-angle, approximately in-plane precession of the free-layer moment. The simulation reproduces the abrupt jump to much lower frequency at the onset of this mode, decreasing frequency with further increases in current (Fig. 3b), and large powers in the harmonics. The maximum simulated microwave powers for this mode in the 0–18 GHz bandwidth are 18 pW mA^{–2} for sample 1 and 75 pW mA^{–2} for sample 2 (differing primarily because of different ΔR_{max} values), whereas the measured maxima are 10 and 90 pW mA^{–2}, respectively. Low-frequency backgrounds in the large-amplitude spectra (for example, Fig. 2d, spectrum 5) might be caused by fluctuations from the large-angle precessional orbit to other modes nearby in energy²⁸. The single-domain simulation does not explain state W in Fig. 2c, but instead for that region it predicts approximately circular out-of-plane precessional modes. These would produce large microwave signals (~ 25 –100 pW mA^{–2}), orders of magnitude larger than the residual signals observed in state W. We suspect that our single-domain approximation may become invalid in regime W owing to dynamical instabilities^{29,30}, so that different regions of the sample may move incoherently, giving total time-dependent resistance changes much smaller than for single-domain motion.

The microwave power generated by the precessing nanomagnet in our devices can be quite significant. For sample 1, the largest peak in the power spectrum has a maximum more than 40 times larger than room-temperature Johnson noise. Nanomagnets driven by spin-polarized currents might therefore serve as nanoscale microwave sources or oscillators, tunable by I and H over a wide frequency range. $\square$

Received 28 May; accepted 28 July 2003; doi:10.1038/nature01967.

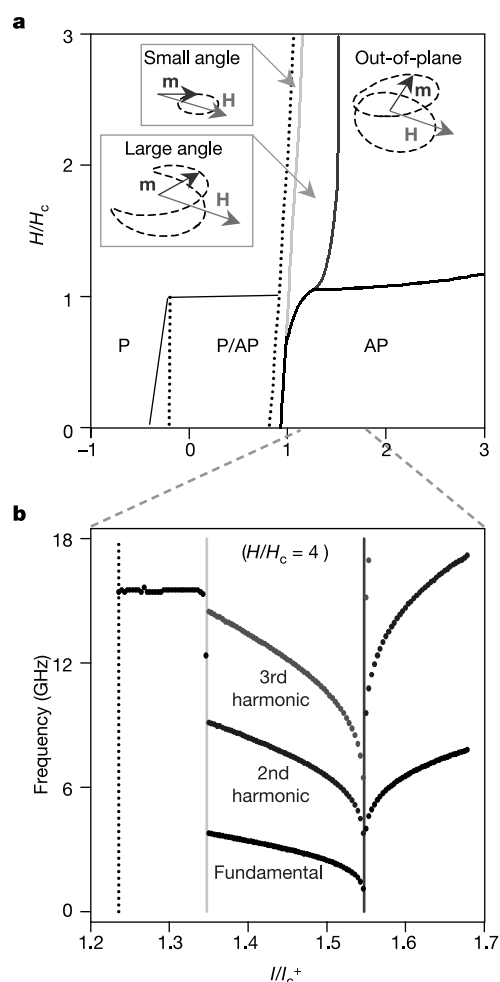


Figure 3 Results of numerical solution of the Landau–Lifshitz–Gilbert equation for a single-domain nanomagnet at zero temperature. The parameters are: $4\pi M_{\text{eff}} = 10$ kOe, $H_{\text{an}} = 500$ Oe, Gilbert damping parameter $\alpha = 0.014$, and effective polarization $P = 0.3$, which produce $H_c = 500$ Oe and $I_c^+ = 2.8$ mA. **a**, Theoretical dynamical stability diagram. The pictures show representative orbits for the free-layer moment vector ($\mathbf{m}$). **b**, Dependence of frequency on current in the simulation for $H = 2$ kOe, including both the fundamental frequency and harmonics in the measurement range.

- Slonczewski, J. C. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* **159**, L1–L7 (1996).
- Berger, L. Emission of spin waves by a magnetic multilayer traversed by a current. *Phys. Rev. B* **54**, 9353–9358 (1996).
- Bazaliy, Y. B., Jones, B. A. & Zhang, S. C. Modification of the Landau–Lifshitz equation in the presence of a spin-polarized current and colossal- and giant-magnetoresistive materials. *Phys. Rev. B* **57**, R3213–R3216 (1998).
- Tsoi, M. *et al.* Excitation of a magnetic multilayer by an electric current. *Phys. Rev. Lett.* **80**, 4281–4284 (1998); erratum *Phys. Rev. Lett.* **81**, 493 (1998).
- Sun, J. Z. Current-driven magnetic switching in manganite trilayer junctions. *J. Magn. Magn. Mater.* **202**, 157–162 (1999).
- Myers, E. B., Ralph, D. C., Katine, J. A., Louie, R. N. & Buhrman, R. A. Current-induced switching of domains in magnetic multilayer devices. *Science* **285**, 867–870 (1999).
- Katine, J. A., Albert, F. J., Buhrman, R. A., Myers, E. B. & Ralph, D. C. Current-driven magnetization reversal and spin-wave excitations in Co/Cu/Co pillars. *Phys. Rev. Lett.* **84**, 4212–4215 (2000).
- Tsoi, M. *et al.* Generation and detection of phase-coherent current-driven magnons in magnetic multilayers. *Nature* **406**, 46–48 (2000).
- Grollier, J. *et al.* Spin-polarized current induced switching in Co/Cu/Co pillars. *Appl. Phys. Lett.* **78**, 3663–3665 (2001).
- Stiles, M. D. & Zangwill, A. Anatomy of spin-transfer torque. *Phys. Rev. B* **65**, 014407 (2002).
- Wegrowe, J.-E. *et al.* Exchange torque and spin transfer between spin polarized current and ferromagnetic layers. *Appl. Phys. Lett.* **80**, 3775–3777 (2002).
- Sun, J. Z., Monsma, D. J., Abraham, D. W., Rooks, M. J. & Koch, R. H. Batch-fabricated spin-injection magnetic switches. *Appl. Phys. Lett.* **81**, 2202–2204 (2002).

13. Bauer, G. E. W., Tserkovnyak, Y., Huertas-Hernando, D. & Brataas, A. Universal angular magnetoresistance and spin torque in ferromagnetic/normal metal hybrids. *Phys. Rev. B* **67**, 094421 (2003).
14. Rippard, W. H., Pufall, M. R. & Silva, T. J. Quantitative studies of spin-momentum-transfer-induced excitations in Co/Cu multilayer films using point-contact spectroscopy. *Appl. Phys. Lett.* **82**, 1260–1262 (2003).
15. Özyilmaz, B. *et al.* Current-induced magnetization reversal in high magnetic fields in Co/Cu/Co nanopillars. *Phys. Rev. Lett.* **91**, 067203 (2003).
16. Urazhdin, S., Birge, N. O., Pratt, W. P. Jr & Bass, J. Current-driven magnetic excitations in permalloy-based multilayer nanopillars. Preprint at (<http://arXiv:cond-mat/0303149>) (2003).
17. Albert, F. J. *The Fabrication and Measurement of Current Perpendicular to the Plane Magnetic Nanostructures for the Study of the Spin Transfer Effect*. PhD dissertation, Cornell Univ. (2003).
18. Baibich, M. N. *et al.* Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices. *Phys. Rev. Lett.* **61**, 2472–2475 (1988).
19. Pozar, D. M. *Microwave Engineering* 2nd edn, 566 (John Wiley & Sons, New York, 1998).
20. Kittel, C. *Introduction to Solid State Physics* 7th edn, 505 (John Wiley & Sons, New York, 1996).
21. Johnson, M. T., Bloemen, P. J. H., den Broeder, F. J. A. & de Vries, J. J. Magnetic anisotropy in metallic multilayers. *Rep. Prog. Phys.* **59**, 1409–1458 (1996).
22. Lee, C. H. *et al.* Magnetic anisotropy in epitaxial Co superlattices. *Phys. Rev. B* **42**, 1066–1069 (1990).
23. Myers, E. B. *et al.* Thermally activated magnetic reversal induced by a spin-polarized current. *Phys. Rev. Lett.* **89**, 196801 (2002).
24. Nazarov, A. V., Cho, H. S., Nowak, J., Stokes, S. & Tabat, N. Tunable ferromagnetic resonance peak in tunneling magnetoresistive sensor structures. *Appl. Phys. Lett.* **81**, 4559–4561 (2002).
25. Sun, J. Z. Spin-current interaction with a monodomain magnetic body: A model study. *Phys. Rev. B* **62**, 570–578 (2000).
26. Bazaliy, Y. B., Jones, B. A. & Zhang, S. C. Towards metallic magnetic memory: How to interpret experimental results on magnetic switching induced by spin-polarized currents. *J. Appl. Phys.* **89**, 6793–6795 (2001).
27. Grolier, J. *et al.* Field dependence of magnetization reversal by spin transfer. *Phys. Rev. B* **67**, 174402 (2003).
28. Li, Z. & Zhang, S. Magnetization dynamics with a spin-transfer torque. *Phys. Rev. B* **68**, 024404 (2003).
29. Suhl, H. The theory of ferromagnetic resonance at high signal powers. *J. Phys. Chem. Solids* **1**, 209–227 (1957).
30. Polianski, M. & Brouwer, P. W. Current-induced transverse spin wave instability in a thin nanomagnet. Preprint at (<http://arXiv:cond-mat/0304069>) (2003).

Acknowledgements We thank K. W. Lehnert, I. Siddiqi and other members of the groups of R. J. Schoelkopf, D. E. Prober and M. H. Devoret for advice about microwave measurements. We acknowledge support from DARPA through Motorola, from the Army Research Office, and from the NSF/NSEC programme through the Cornell Center for Nanoscale Systems. We also acknowledge use of the NSF-supported Cornell Nanofabrication Facility/NNUN.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.C.R. (ralph@cmmr.cornell.edu).

Video-speed electronic paper based on electrowetting

Robert A. Hayes & B. J. Feenstra

Philips Research Eindhoven, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

In recent years, a number of different technologies have been proposed for use in reflective displays^{1–3}. One of the most appealing applications of a reflective display is electronic paper, which combines the desirable viewing characteristics of conventional printed paper with the ability to manipulate the displayed information electronically. Electronic paper based on the electrophoretic motion of particles inside small capsules has been demonstrated¹ and commercialized; but the response speed of such a system is rather slow, limited by the velocity of the particles. Recently, we have demonstrated that electrowetting is an attractive technology for the rapid manipulation of liquids on a micrometre scale⁴. Here we show that electrowetting can also be used to form the basis of a reflective display that is significantly faster than electrophoretic displays, so that video content can be displayed. Our display principle utilizes the voltage-controlled movement of a coloured oil film adjacent to a white substrate.

The reflectivity and contrast of our system approach those of paper. In addition, we demonstrate a colour concept, which is intrinsically four times brighter than reflective liquid-crystal displays⁵ and twice as bright as other emerging technologies^{1–3}. The principle of microfluidic motion at low voltages is applicable in a wide range of electro-optic devices.

Microfluidic movement based on electrowetting^{4,6,7}—where a voltage difference between a hydrophobic solid and a liquid causes a change in wettability—is being used for an increasing number of applications. These include pixelated optical filters⁴, adaptive lenses⁸ and lab-on-a-chip⁹. Electrowetting has several very attractive features for use in micrometre- to millimetre-sized systems: low power consumption, fast response speed and scalability. In addition, when a fluoropolymer coating with low contact-angle hysteresis is used, a high degree of reversibility can be obtained⁴. However, in most of the applications reported a thick insulator is used, giving rise to high switching voltages. By improving the processing of hydrophobic insulating materials we managed to lower the drive voltages dramatically^{10,11}, opening up a much broader application area.

For the electrowetting display principle, the focus is on the movement of a confined water–oil interface (Fig. 1). In equilibrium, a coloured oil film lies naturally between the water and the hydrophobic insulator coating of an electrode, because

$$\gamma_{o,w} + \gamma_{o,i} < \gamma_{w,i} \quad (1)$$

where γ is the interfacial tension, and the subscripts denote the oil, water and insulator, respectively. Owing to the dominance of interfacial over gravitational forces in small systems (<2 mm), such an oil film is continuous and stable in all orientations. However, when a voltage V is applied between the substrate electrode and the water, an electrostatic term ($\sim 0.5CV^2$, where C is the parallel-plate capacitance) is added to the energy balance, and the stacked state is no longer energetically favourable (Fig. 1b). The system can lower its energy by moving the water into contact with the insulator, thereby displacing the oil.

The balance between electrical and capillary forces determines how far the oil is moved to the side. Hence the optical properties of the stack, when viewed from above, can be continuously and reversibly tuned between a coloured off-state and a transparent on-state, assuming that the pixel is sufficiently small that a viewer

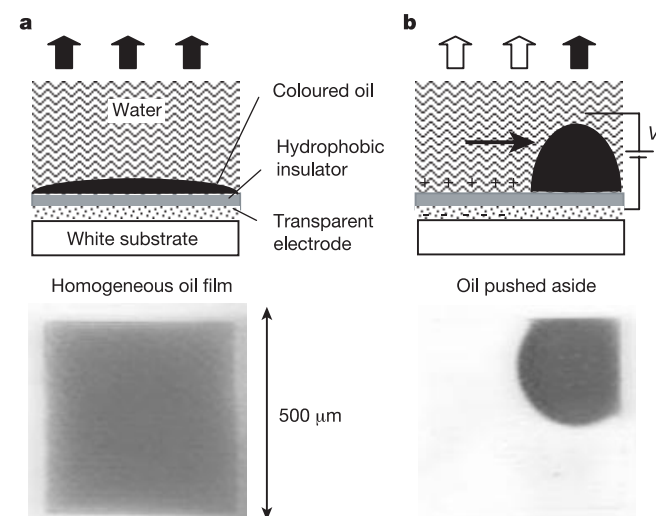


Figure 1 Electrowetting display principle. **a**, No voltage applied, therefore a coloured homogeneous oil film is present. **b**, d.c. voltage applied, causing the oil film to contract. Top row, diagrams; bottom row, photographs. The photographs show typical oil motion obtained with an homogeneous pixel electrode.

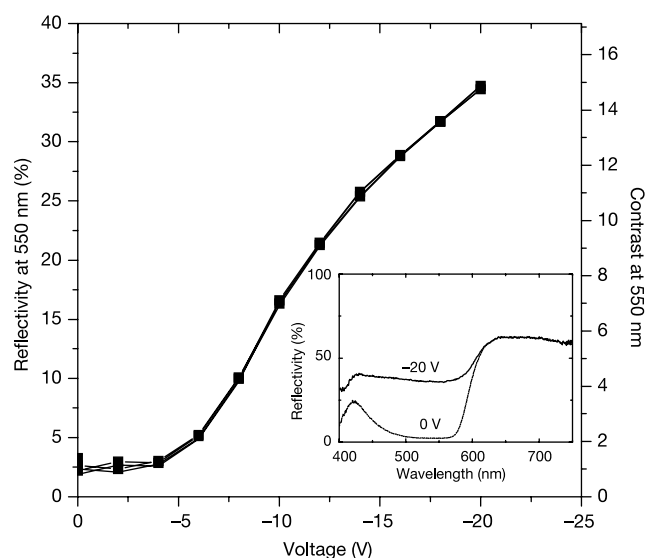


Figure 2 Electro-optic behaviour of electrowetting display pixels. Reflectivity and contrast as a function of d.c. voltage for a $500 \times 500 \mu\text{m}^2$ pixel with a $15\text{-}\mu\text{m}$ -thick magenta oil layer and a $0.8\text{-}\mu\text{m}$ -thick fluoropolymer insulator. Inset, full wavelength response.

experiences the average optical response. Alternatively, if a diffusely reflective (white) surface is placed under the switchable element, a simple high-brightness/high-contrast optical switch is obtained that can be used as the basis for a reflective display. It has been previously proposed¹² that electrowetting could be used as a display principle by driving a refractive index matched liquid into a three-dimensional porous network to provide the optical modulation from white to transparent. Optical performance aside, applying the metallic and insulating layers required for reversible electrowetting in such a porous network renders this principle highly impractical. In contrast, the microfluidic motion demonstrated here is two-dimensional, which can be practically realized using the currently available (and reliable) electrowetting device materials over large display areas.

In our system, the white reflector is integrated into the optical stack by coating a white polymer foil with a thin, patterned (15 nm indium tin oxide, ITO) electrode layer and a fluoropolymer insulator. The moving oil film is now separated from the white substrate by just the thickness of the electrode/insulator combination (typically $<1\text{ }\mu\text{m}$), resulting in a very efficient recycling of ambient light. Oils of any required colour are formulated by dissolving non-polar dyes in alkanes (typically $\text{C}_{10}\text{--C}_{16}$). As a large variety of oils and polar liquids can be used, we can adapt our system to the temperature requirements. To contain the oil films on a pixel resolution, we use a thin film fabrication procedure. A black or transparent polymer sheet, typically $50\text{ }\mu\text{m}$ thick, is laser cut to provide the required pixel sizes and patterns. It is then glued onto the insulator-covered substrate with a low viscosity two-part epoxy. Liquids are dosed and the test cells are sealed with an ITO-covered glass slide.

The electro-optic characteristics of a single electrowetting pixel measured at 0° with diffuse illumination are shown in Fig. 2. A small threshold voltage is observed before displacement of the oil film commences. Very little hysteresis in fluid motion is observed. The intermediate optical states are stable, implying that analogue, voltage-controlled grey scales can be realized. At a voltage of about -20 V , the oil film has contracted dramatically and in excess of 70% of the white substrate is exposed, resulting in a reflectivity of about 35% (left-hand axis in Fig. 2). At higher voltages, where the electro-optic curve plateaus, up to 90% of the white substrate becomes exposed. The contraction of the oil results in a maximum contrast of

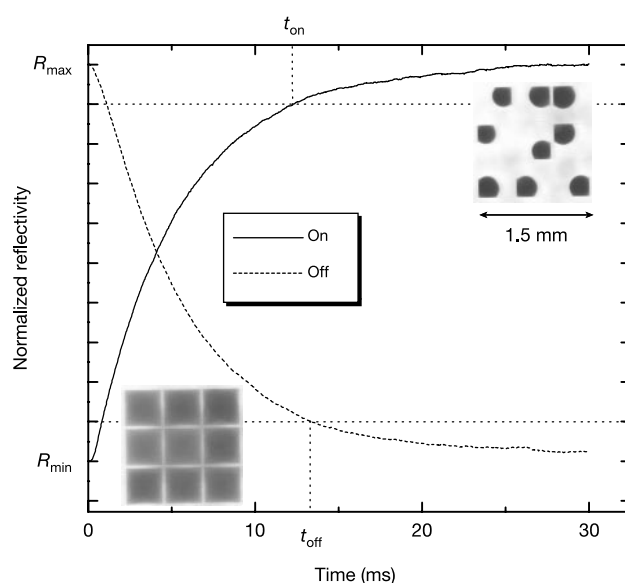


Figure 3 Electrowetting pixel kinetics. Temporal behaviour of a $250 \times 250 \mu\text{m}^2$ pixel, demonstrating the video-speed response. The oil film thickness is $15\text{ }\mu\text{m}$ and the insulator thickness is $0.8\text{ }\mu\text{m}$. The on and off response times are 12 and 13 ms respectively. The reflectivity is shown in normalized units, which does not affect the time axis. t_{on} and t_{off} are defined as the time it takes to complete 90% of the optical modulation. Insets, photographs showing the corresponding optical state for a 3×3 array of $500 \times 500 \mu\text{m}^2$ pixels. With a homogeneous pixel electrode, the observed oil motion is reproducible for a given pixel but is variable between pixels. Motion to a specific position in a pixel array can be realized by using an inhomogeneous pixel electrode (Supplementary Information).

about 15. We determined the contrast (the reflectivity ratio with and without voltage) and the reflectivity at the centre of the absorption band (550 nm), because we used a coloured dye that absorbs only part of the visible spectrum. The full wavelength response (Fig. 2 inset) is indicated for 0 and -20 V . The reflectivity R and contrast obtained are comparable to the optical properties of electrophoretic displays ($R \approx 40\%$ and contrast ~ 11)¹³, and are approaching the optical performance of paper ($R \approx 60\%$ and contrast ~ 15).

The main pixel parameters that influence the electro-optical performance are the thickness of the oil film and the thickness of the fluoropolymer insulator. As the oil film becomes thicker, the electro-optic curve is shifted to higher field strength. As a result, it is best to work with an oil film thickness that is sufficiently thin to give low operating voltage, while sufficiently thick to ensure a satisfactory optical activity (typically around $10\text{ }\mu\text{m}$, that is, nanolitre volumes of oil). The thickness of the insulator determines the necessary field strength as soon as the oil has moved (partly) aside. Hence the insulator should be as thin as possible, while still having sufficient dielectric strength¹⁰.

In small pixels, the motion of the oil film is sufficiently fast to show video content (Supplementary Information). The optically measured response speed in a $250 \times 250 \mu\text{m}^2$ pixel (equivalent to 100 pixels per inch) is shown in Fig. 3. Both the on-switch and the off-switch show a response time close to 10 ms. The on-switch is voltage driven, while the off-switch relies solely on capillary forces to reform the oil film. We have recently commenced the evaluation of larger arrays of electrowetting pixels, and introduced individual pixel addressing (Supplementary Information). Sealed pixel arrays exhibit no deterioration in electro-optic performance after several million switches. The individual pixel addressing in larger-size displays can be provided by active matrix technology.

One of the major drawbacks of current reflective display technologies is their low colour-reflectivity, mostly because of the use of RGB (red-green-blue) colour segmentation¹⁻³. In Fig. 4a we show a

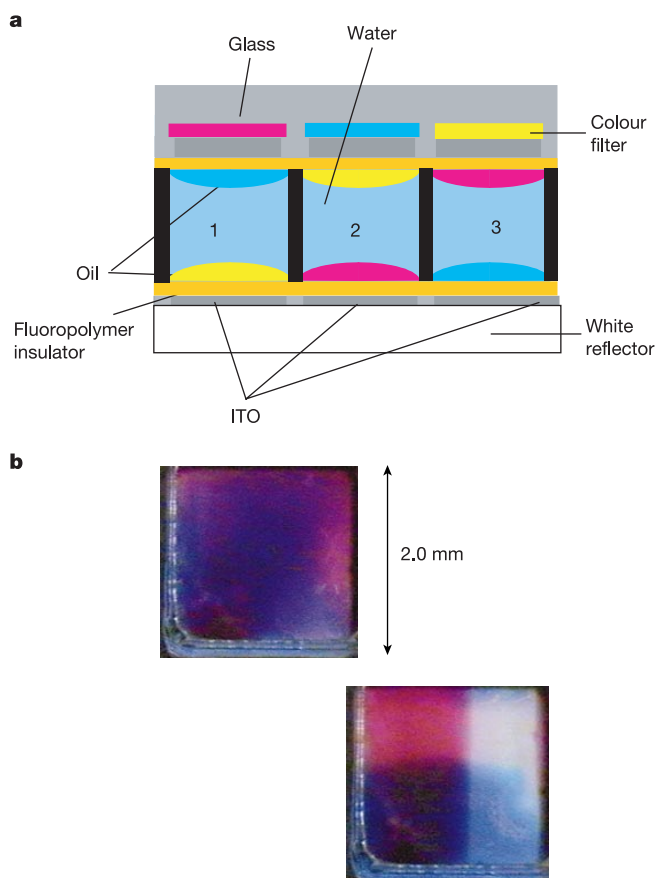


Figure 4 The high-brightness colour electrowetting display principle. **a**, Diagram showing a single subpixelated display pixel with two active layers. **b**, Top view of working cell containing magenta (top) and cyan (bottom) oil films separated by water.

colour concept for an electrowetting display that has a reflectivity that is four times higher than that of a reflective colour liquid-crystal display (LCD) and twice as high as other emerging technologies. In this case, a single display pixel would comprise three subpixels. In each of the subpixels a second oil layer is added, adjacent to the top plate. With a hydrophilic pixel wall, the sandwich structure with two oil layers is the stable equilibrium state. Each oil layer can be switched independently and reversibly by applying a voltage difference between the water and the top or bottom electrode. As for the single layer system, intermediate grey scales can be realized, for each of the two independent switches in a subpixel separately. By using a colour filter, subtractive colour dyes and alternating the colour sequence in the subpixels, we obtain a pixel that has an intrinsic reflectivity that is four times higher than an LCD (67% versus 17%). The oil films can be separated by less than 100 μm , resulting in negligible parallax and a good viewing angle. A demonstration of the independent switching of two layers in a single cell is given in Fig. 4b. Here, the electrodes cover only half of the top and bottom plate, orthogonally oriented with respect to one another. The top layer is magenta and the bottom layer is cyan. As a result, a different optical response is obtained in the bottom left (two layers present), the upper left and bottom right (one layer) and the upper right corner (both layers absent).

We have shown that the microfluidic motion of coloured oil films sandwiched between a solid hydrophobic insulator and water can be precisely and reversibly controlled using a d.c. voltage. We have used this phenomenon to demonstrate electrowetting as a reflective display principle that has very attractive electro-optical characteristics: a reflectivity of over 35%, a contrast of 15 and analogue grey scales. Its high switching speed and a straightforward path to a high-

brightness full colour display set it apart from other emergent electronic-paper technologies. □

Received 16 May; accepted 11 August 2003; doi:10.1038/nature01988.

1. Comiskey, B., Albert, J. D., Yoshizawa, H. & Jacobson, J. An electrophoretic ink for all-printed reflective electronic displays. *Nature* **394**, 253–255 (1998).
2. Sheridan, N. K. *et al.* in *Proc. IDRC97* (ed. Morreale, J.) L82–L85 (Society for Information Display, Toronto, 1997).
3. Podojil, G. M. *et al.* in *SID98 Digest* (ed. Morreale, J.) 51 (Society for Information Display, Anaheim, CA, 1998).
4. Prins, M. W. J., Welters, W. J. J. & Weekamp, J. W. Fluid control in multichannel structures by electrocapillary pressure. *Science* **291**, 277–279 (2001).
5. Grupp, J. in *Eurodisplay Digest 2002* 35–38 (Le Club Visu, SID-France, Nice, 2002).
6. Welters, W. J. J. & Fokkink, L. G. J. Fast electrically switchable capillary effects. *Langmuir* **14**, 1535–1538 (1998).
7. Quilliet, C. & Berge, B. Electrowetting: A recent outbreak. *Curr. Opin. Colloid Interface Sci.* **6**, 34–39 (2001).
8. Berge, B. & Peseux, J. Variable focus lens controlled by an external voltage: An application of electrowetting. *Eur. Phys. J. E* **3**, 159–163 (2000).
9. Pollack, M. G., Fair, R. B. & Shenderov, A. Electrowetting-based actuation of liquid microdroplets for microfluidic applications. *Appl. Phys. Lett.* **77**, 1725–1726 (2000).
10. Seyrat, E. & Hayes, R. A. Amorphous fluoropolymers as insulators for reversible low-voltage electrowetting. *J. Appl. Phys.* **90**, 1383–1386 (2001).
11. Moon, H., Sung, K. C., Garrell, R. L. & Kim, C. J. Low voltage electrowetting-on-dielectric. *J. Appl. Phys.* **92**, 4080–4087 (2002).
12. Beni, G. & Hackwood, S. Electrowetting displays. *Appl. Phys. Lett.* **38**, 207–209 (1981).
13. Ritter, J. in *IDW01 Digest* 343 (Society for Information Display, Nagoya, 2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We acknowledge the infrastructural support provided by the Devices, Technology & Services Department, our project colleagues and our team of students (M. Joulaud, J. Guiberteau, R. Massard, E. Morks, A.-S. Dupont, K. Girard, C. Maufrains and P. Jaulneau).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.A.H. (rob.hayes@philips.com) or B.J.F. (johan.feenstra@philips.com).

A combustion-free methodology for synthesizing zeolites and zeolite-like materials

Hyunjoo Lee¹, Stacey I. Zones² & Mark E. Davis¹

¹Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

²ChevronTexaco Energy and Research Center, 100 Chevron Way, Richmond, California 94802, USA

Zeolites are mainly used for the adsorption and separation of ions and small molecules, and as heterogeneous catalysts. More recently, these materials are receiving attention in other applications, such as medical diagnosis and as components in electronic devices¹. Modern synthetic methodologies for preparing zeolites and zeolite-like materials typically involve the use of organic molecules that direct the assembly pathway and ultimately fill the pore space^{2–6}. Removal of these enclathrated species normally requires high temperature combustion that destroys this high cost component, and the associated energy release in combination with the formed water can be extremely detrimental to the inorganic structure⁷. Here we report a synthetic methodology that avoids these difficulties by creating organic structure-directing agents (SDAs) that can be disassembled within the zeolite pore space to allow removal of their fragments for possible use again by reassembly. The methodology is shown for the synthesis of zeolite ZSM-5 using a SDA that

contains a cyclic ketal group that is removed from the SDA while it is inside the zeolite without destruction of the inorganic framework. This approach should be applicable to the synthesis of a wide variety of inorganic and organometallic structures.

The elimination of high temperature treatments for SDA removal from crystalline structures is very desirable for many reasons in addition to the loss of the expensive SDA. For example, zeolite films that are used as molecular sieving membranes are susceptible to cracking by high temperature treatments for SDA removal because of mechanical stresses placed on the membrane by thermal expansion mismatches with supporting substrates. Newer, molecular sieve low dielectric components⁸ may also benefit from the ability to remove the 'guest' organic molecule using only moderate heating, rather than combustion. This is because they require air to fill the microporous space in order to achieve their desired properties and will be components of devices that may not be compatible with high temperature processing steps¹.

Ordered, mesoporous materials can be prepared using organic components such as surfactants that can form organized aggregates that contain large numbers of molecules that ultimately fill the pore space of the as-synthesized solids^{9,10}. Unlike crystalline microporous materials, the mesoporous solids allow for the extraction of the organic structure-directing components^{11–13}. This is due to the fact that the individual molecules that form the assembled structure-directing components are held together by weak forces that are easily disrupted and are each sufficiently small to be removed

through the relatively large mesopores. Additionally, the interaction energies between the organic and inorganic fractions in the ordered, mesoporous materials are not as large as with microporous materials, where interactions between the SDA molecules and the framework can be quite strong. For example, interaction enthalpies between the silica framework and the organic SDA have been measured to be as large as -181 ± 21 kJ per mol SDA¹⁴. The ability to remove and recycle the organic components allows for low-cost preparation of ordered, mesoporous materials.

Our synthetic strategy for microporous crystalline molecular sieves (Fig. 1a) eliminates any high temperature processing steps and the destruction of the SDA. Conceptually, the strategy is to assemble a SDA from at least two components using covalent bonds and/or non-covalent interactions that are able to survive the conditions for assembly of the zeolite, and yet be reversed inside the microporous void space. The fragments formed from the SDA in the zeolite can then be removed from the inorganic framework and be recombined for use again. This procedure differs from that employed with mesoporous materials in that the number of components used to create the SDA is small—for example, less than five—and the SDA is a homogeneous, well-defined entity.

We demonstrate our synthetic methodology with a new SDA that contains a cyclic ketal and use it to assemble the zeolite ZSM-5 (Fig. 1b). Commercially available 1,4-dioxo-8-azaspiro[4,5]decane was quaternized using methyl iodide to give **1** (see below).

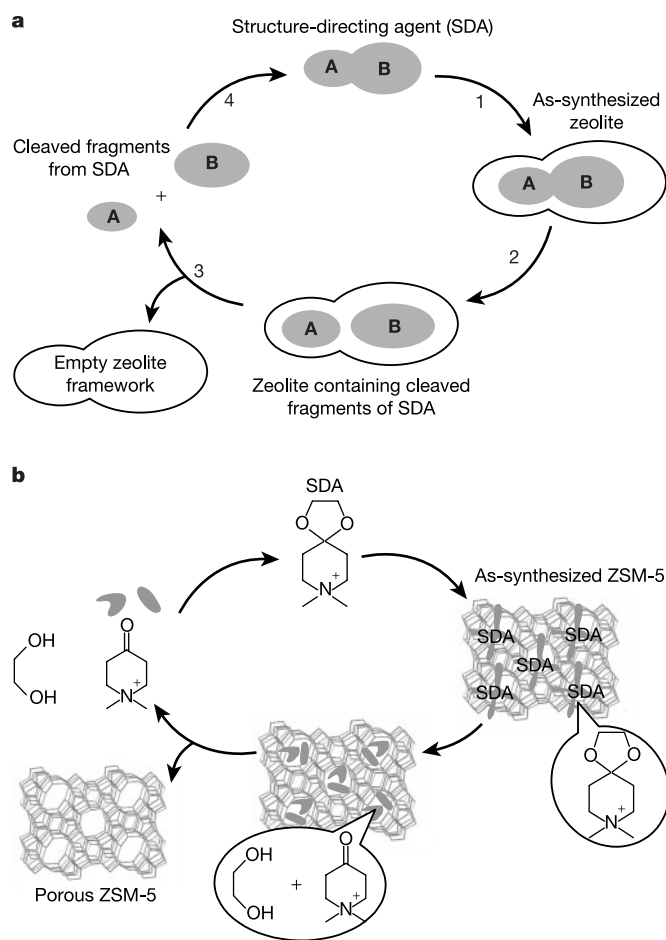
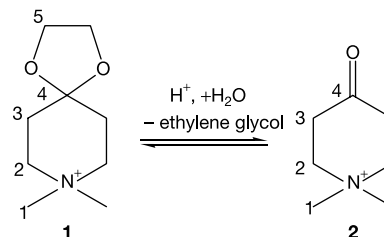


Figure 1 Schematic representations of synthetic methodology. **a**, Generalized scheme. Step 1, assemble the SDA with silica precursor, H_2O , alkali metal ions, and so on, for zeolite synthesis. Step 2, cleave the organic molecules inside the zeolite pores. Step 3, remove the fragments. Step 4, recombine the fragments into the original SDA molecule. **b**, Specific example using **1** to give ZSM-5.



Using **1** in its hydroxide form, the following reaction composition gave ZSM-5 (submicrometre crystals observed in scanning electron microscopy: see Supplementary Fig. 1); 0.033 **1**: 0.238 KOH: 0.056 $Al(OH)_3$: 42.23 H_2O : 1.0 SiO_2 (Fig. 2a). In the absence of the SDA, this reaction mixture does not yield ZSM-5. As-synthesized ZSM-5 has an elemental composition of 40.27 wt% Si, 2.31 wt% Al, 2.25 wt% K, 3.45 wt% C, 0.97 wt% H, 0.45 wt% N. The SiO_2/Al_2O_3 ratio was estimated from these analyses to be 33.5.

We chose to use a cyclic ketal to provide a SDA that would remain intact at zeolite synthesis conditions (high pH) and be cleavable at conditions that would not destroy the assembled zeolite (low pH). **1** was expected to be cleaved into **2** and ethylene glycol under acidic condition. Figure 3 shows the ^{13}C cross-polarization magic angle spinning (CP MAS) NMR spectra from the ZSM-5 after various treatments. The spectrum in Fig. 3a shows that the as-synthesized material contains intact **1**. When the ZSM-5 was treated with 1 M HCl solution at 80 °C for 20 h, the ^{13}C CP MAS NMR spectrum obtained (Fig. 2b) is consistent with the presence of the ketone fragment, as illustrated above. The carbonyl of **2** was also detected by infrared spectroscopy ($1,741\text{ cm}^{-1}$), and the weight per cent of the organic components as assessed by thermogravimetric weight losses at temperatures between 200 and 700 °C decreased from 6.2 wt% to 4.6 wt% (consistent with the loss of ethylene glycol). The HCl solution after the cleavage reaction was collected, concentrated, and then analysed by ^{13}C NMR. Only ethylene glycol was detected at 62.8 p.p.m. The powder X-ray diffraction pattern of the treated sample reveals that the ZSM-5 remains intact after HCl treatment (Fig. 2b). Alternatively, a gas-phase cleavage reaction was performed using H_2O -saturated gaseous HCl at 120 °C for 3 h. The

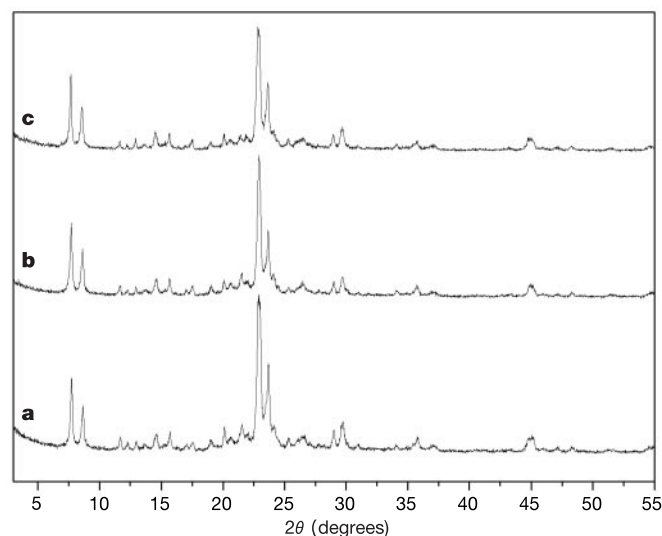


Figure 2 Powder X-ray diffraction patterns. **a**, As-synthesized ZSM-5. **b**, After treatment with 1 M HCl solution. **c**, After ion exchange with 0.01 M NaOH + 1 M NaCl solution. The ZSM-5 shows no structural degradation after the various treatments.

^{13}C CP MAS NMR spectrum of the treated ZSM-5 was the same as that shown in Fig. 3b except that ethylene glycol was also detected. These results showed that **1** can be cleaved into the desired pieces inside the zeolite pore space.

While ethylene glycol can be extracted easily, **2** remains inside the pores because of strong ionic interaction with the anionic framework. Ion exchange was used to remove these positively charged organic fragments. Exposure to a mixture of 0.01 M NaOH and 1 M NaCl at 100 °C for 72 h removed **2** (Fig. 3c, and no carbonyl peak observed in infrared). Additionally, the ion-exchanged ZSM-5 now has porosity (N_2 adsorption capacity of 0.14 cm^3 liquid N_2 per g dry ZSM-5) that was not present before this treatment (as-synthesized ZSM-5 showed no microporosity, while the calcined ZSM-5 gave $0.15 \text{ cm}^3 \text{ g}^{-1}$). When the as-synthesized solid was not treated with HCl solution, the ion-exchange step did not cause the removal of **1**. Thus, the fragmentation of **1** is essential for its removal from the zeolite pores. **2** was detected in the solution after the ion exchange by electrospray mass spectrometry. After NaOH/NaCl treatment, the powder X-ray diffraction data show no loss in structural integrity of the ZSM-5 (Fig. 2c). Additionally, the ^{27}Al NMR spectrum from the NaOH/NaCl treated material (see Supplementary Fig. 2) contains only the resonance for framework aluminium at 55.1 p.p.m. No extra-framework aluminium is detected. For numerous synthetic preparations that have been treated by either the liquid phase HCl method or the vapour phase HCl method, the measured $\text{SiO}_2/\text{Al}_2\text{O}_3$ ratios are 30 ± 4 . Thus, within the error reported for the $\text{SiO}_2/\text{Al}_2\text{O}_3$ ratio, the composition of the samples did not vary over the treatment regimes. The catalytic activity of this ZSM-5 was evaluated using the conversion of methanol to higher hydrocarbons (MTG) as the test reaction. The product distribution of individual hydrocarbons was the same (within experimental errors) in distribution to that obtained from a commercially available sample of ZSM-5 (see Supplementary Table 1). Additionally, the constraint index (CI) that measures the relative cracking rates of *n*-hexane and 3-methylpentane¹⁵ was determined for the ammonium forms of our synthesized ZSM-5 that was calcined or processed through the organic removal procedure outlined above. This test reaction should be more discerning than the MTG reaction in terms of differences in acid strength and number of sites when comparing zeolite samples. The samples show similar rates and CIs (see Supplementary Table 2). Additionally, these data compare well to those obtained from a commercial sample of ZSM-5^{15,16}.

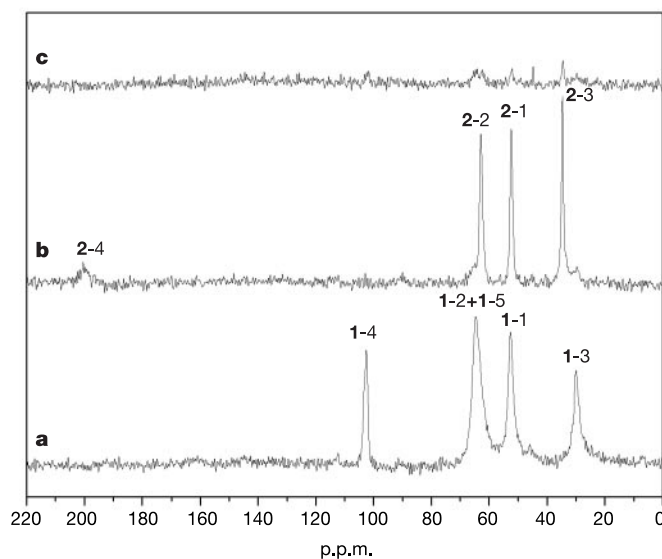


Figure 3 ^{13}C CP MAS NMR spectra. **a**, Intact **1** inside as-synthesized ZSM-5. **b**, After cleavage of **1** inside ZSM-5 pores using 1 M HCl solution. **c**, After ion exchange with 0.01 M NaOH + 1 M NaCl solution. The spinning rate was 4 kHz and contact time for cross-polarization was 1 ms for all cases. The notation of **1-1** denotes the peak corresponding to the carbon 1 of the organic molecule **1** (see text). The ^{13}C CP MAS NMR data show that **1** was cleaved into its desired fragments by acid treatment, and successfully removed after ion exchange.

Here we have provided the essential results necessary for the demonstration of our proposed synthetic methodology. The recombination of the fragments of **1** is not illustrated, but this reaction is well-known. We have successfully synthesized other zeolites (ZSM-11, ZSM-12) using this generalized methodology. In addition to ketals like those illustrated here, acetals and *ortho*-ester functionalities have the appropriate properties (hydrolytically acid labile and base stable)¹⁷ for use in the outlined synthesis methodology. We are currently applying this methodology to the preparation of larger SDAs, and testing these organics in a broad selection of zeolite synthesis conditions. □

Methods

Preparation of SDA (**1**)

4.00 g of 1,4-dioxo-8-azaspiro[4,5]decane (98%, Aldrich), 8.01 g of tributylamine (99%, Aldrich), and 30 ml of MeOH (Burdick & Jackson) were mixed in a flask. 12.20 g of iodomethane (99.5%, Aldrich) were added dropwise over a period of 10 min. The mixture was refluxed for 5 days at room temperature. Yellow solids were produced. After adding ethyl ether to the mixture, the produced solids were filtered and washed with ethyl ether. The solids were recrystallized from hot acetone/MeOH. Iodine salts were converted to the corresponding hydroxide form in 90.2% yield using Bio-Rad AG1-X8 anion exchange resin.

Preparation of as-synthesized zeolite

0.2 g of potassium hydroxide (Aldrich) and 0.083 g of aluminium hydroxide (Reheis F-2000) were added to the mixture of 0.094 g of **1** and 11.4 g of water, and stirred to obtain a clear solution. 0.9 g of silica (CabOSil M5) were added to the solution and the mixture was stirred for 2 h to prepare a homogeneous gel. The resulting mixture was charged into a rotating Teflon-lined autoclave (100 r.p.m.) and heated at 175 °C for 6 days. After crystallization, the autoclave was cooled to room temperature. The solid product was collected by filtration, repeatedly washed with deionized water and finally dried overnight.

Characterizations

^{13}C CP MAS NMR measurement was performed with a Bruker Avance 200 MHz spectrometer. Solution ^{13}C NMR spectrum was obtained with a Varian Mercury 300 MHz spectrometer. Powder X-ray diffraction patterns were collected on a Scintag XDS 2000 diffractometer using CuK α radiation at the rate of $2.5^\circ \text{ min}^{-1}$. Elemental analysis was obtained by Quantitative Technologies Inc. Infrared spectroscopy was performed on a Nicolet Nexus 470 FT-IR. Thermogravimetric analysis was performed on a NETZSCH STA 449C analyser. Nitrogen adsorption capacity was measured using an Omnisorp 100 sorption apparatus. Mass spectrum was measured with a Perkin Elmer/Sciex API 365 electrospray mass spectrometer.

Received 27 March; accepted 5 August 2003; doi:10.1038/nature01980.

1. Davis, M. E. Ordered porous materials for emerging applications. *Nature* **417**, 813–821 (2002).
2. Freyhardt, C. C., Tsapatsis, M., Lobo, R. F., Balkus, K. J. & Davis, M. E. A high-silica zeolite with a 14-tetrahedral-atom pore opening. *Nature* **381**, 295–298 (1996).
3. Wagner, P. et al. CIT-5: A high-silica zeolite with 14-ring pores. *Chem. Commun.* 2179–2180 (1997).
4. Wagner, P. et al. Guest/host relationships in the synthesis of the novel cage-based zeolites SSZ-35, SSZ-36, and SSZ-39. *J. Am. Chem. Soc.* **122**, 263–273 (2000).
5. Wright, P. A. et al. Cation-directed syntheses of novel zeolite-like metalloaluminophosphates STA-6 and STA-7 in the presence of azamacrocyclic templates. *J. Chem. Soc. Dalton Trans.* 1243–1248 (2000).
6. Corma, A., Diaz-Cabanas, M. J., Martinez-Triguero, J., Rey, F. & Rius, J. A large-cavity zeolite with wide pore windows and potential as an oil refining catalyst. *Nature* **418**, 514–517 (2002).
7. Kuehl, G. H. & Timken, H. K. C. Acid sites in zeolite beta: Effects of ammonium exchange and steaming. *Micropor. Mesopor. Mater.* **35–36**, 521–532 (2000).
8. Wang, Z., Wang, H., Mitra, A., Huang, L. & Yan, Y. Pure-silica zeolite low-k dielectric thin films. *Adv. Mater.* **13**, 746–749 (2001).
9. Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *Nature* **359**, 710–712 (1992).
10. Corma, A. From microporous to mesoporous molecular sieve materials and their use in catalysis. *Chem. Rev.* **97**, 2373–2419 (1997).
11. Whitehurst, D. D. Method to recover organic templates from freshly synthesized molecular sieves. US Patent 5,143,879 (1992).
12. Chen, C. Y., Li, H. X. & Davis, M. E. Studies on mesoporous materials I. Synthesis and characterization of MCM-41. *Micropor. Mater.* **2**, 17–26 (1993).
13. Tanev, P. T. & Pinnavaia, T. J. A neutral templating route to mesoporous molecular sieves. *Science* **267**, 865–867 (1995).
14. Piccione, P. M., Yang, S., Navrotsky, A. & Davis, M. E. Thermodynamics of pure-silica molecular sieve synthesis. *J. Phys. Chem. B* **106**, 3629–3638 (2002).
15. Frillette, V. J., Haag, W. O. & Lago, R. M. Catalysis by crystalline aluminosilicates: Characterization of intermediate pore-size zeolites by the “constraint index”. *J. Catal.* **67**, 218–222 (1981).
16. Zones, S. I. & Harris, T. V. The constraint index test revisited: Anomalies based upon new zeolite structure types. *Micropor. Mesopor. Mater.* **35–36**, 31–46 (2000).
17. Srinivasachar, K., David, M. & Neville, J. New protein cross-linking reagents that are cleaved by mild acid. *Biochemistry* **28**, 2501–2509 (1989).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. J. Hwang for assistance with NMR measurements, and L. T. Yuen and B. Lee for performing the catalytic testing. This work was supported by the ChevronTexaco Research and Technology Company.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.E.D. (mdavis@chem.caltech.edu).

An Arctic mammal fauna from the Early Pliocene of North America

Richard H. Tedford¹ & C. Richard Harington²

¹Division of Paleontology, American Museum of Natural History, New York, New York 10024, USA

²Canadian Museum of Nature (Paleobiology), Ottawa, Ontario K1P 6P4, Canada

A peat deposit on Ellesmere Island¹, Nunavut, Canada, allows a unique glimpse of the Early Pliocene terrestrial biota north of the Arctic Circle. The peat accumulated in a beaver pond surrounded by boreal larch forest near regional tree line² in coastal hills close to the Arctic Ocean. The ecological affinities of the plant and beetle remains³ contained in the peat indicate that winter temperatures on Ellesmere Island were nearly 15 °C higher and summer temperatures 10 °C higher than they are today. Here we show that the mammalian remains buried in the peat represent mainly taxa of Eurasiatic zoogeographic and phyletic affinities, including the first North American occurrence of a meline badger (*Arctomeles*). This deposit contains direct evidence of the composition of an Early Pliocene (4–5 million years ago) arctic mammalian fauna during an active period of interchange between Asia and North America.

The Early Pliocene assemblage comes from outcrops in Strath-

cona Fiord on Ellesmere Island at 78° 33' N, 82° 22' W (Fig. 1), where the fossiliferous peat is intercalated in a sequence of sands and gravels 10–40 m thick that cap the eroded surface of the Eocene Eureka Sound Group. This high-level alluvium (approximately 300 m above sea level) was deposited before the excavation of the present valleys and fiords. The peat contains well-preserved plant remains¹, with rooted larch trunks up to 3 m tall and many beaver-cut branches and saplings⁴. A possible unconformity of the lower part of the deposit, and floral differences from the overlying peat, suggest that the high-level alluvium may span a significant interval of time. However the beaver pond deposit represents a coeval association of remains.

The Beaver Pond flora includes macrofossils of bryophytes and vascular plants, the former mostly extant species. The latter contains some extinct forms including a larch (the fossil *Larix groenlandii* rather than living tree-line species, *L. laricina*), spruce (*Picea*) and a pine close to the living Japanese stone pine (*Pinus pumila*)². Alder (*Alnus*) and birch (*Betula*) grew near the pond⁴. This flora was richer than that of present-day North American boreal forests especially in being augmented by taxa now confined to Eurasia. The beaver pond was surrounded by open larch woodland near regional tree line more than 2,000 km north of the present tree line.

A diverse beetle (Coleoptera) fauna of 16 extant species is also contained in the peat. These were subjected to the mutual climatic range method of palaeotemperature analysis³ to estimate mean summer and winter temperatures for the site. The results, consistent with that from the associated flora, indicated that the site was 10 °C warmer in summer and 15 °C warmer in winter than today, similar to modern values for Labrador.

The mammal fauna⁴ of the Beaver Pond site is more diverse than that of the modern Arctic tree line and is dominated by Eurasiatic taxa. Only the castoroidine beaver *Dipoides* (cf. *D. intermedius*), the archaeolagine rabbit *Hypolagus* (cf. *H. vetus*) and a small canine species belonging to the extinct genus *Eucyon* have long phyletic histories in mid-continent North America before their appearances in eastern Asia in the Late Miocene to Early Pliocene. These records confirm the adaptation of such taxa to the high-latitude environments of the Bering land bridge at the time of their entrance into Asia.

Other elements of the fauna have centres of diversity and phyletic histories in Eurasia: a neomyine shrew, *Arctisorex polaris*⁵, a microtine-like cricetid similar to *Microtodon* or *Promimomys*⁶, and an ursine bear, *Ursus abstrusus*⁷, more primitive than those known in Eurasia. This bear appears as an immigrant in mid-latitude deposits of North America dating from 3.5–5.0 million years (Myr) ago^{7,8}. Three mustelids are present: a large wolverine (cf. *Plesiogulo*), a fisher (*Martes*) and, surprisingly, a meline badger (*Arctomeles sotnikovae* sp. nov.). Living Eurasian badgers (Mustelidae, subfamily Melinae) belong to two distinctive genera, *Meles* and the hog-badger *Arctonyx*, that occupy temperate and tropical Eurasia, respectively, overlapping only in eastern China (Fig. 1). These genera have fossil records that extend into the Early Pliocene for *Meles* and medial Pleistocene for *Arctonyx*, limited to the same geographic ranges as their living species. The extinct badger *Arctomeles* contains several closely similar fossil species and appears to be the sister taxon of *Arctonyx*⁹. Species of *Arctomeles* are confined to the Early Pliocene and range co-extensively with *Meles* across temperate Eurasia (Fig. 1). The Beaver Pond form (see Methods) is the most distinctive of these in both its morphology and its geographic occurrence.

Also present is a tridactyl horse belonging to the Asiatic genus *Plesiohipparion*¹⁰, and a deerlet similar to the living musk deer *Moschus*, which lives in Siberia today. Most of these Eurasian taxa have their earliest joint occurrence in the latest Miocene and Early Pliocene deposits of the Yushe Basin¹¹, northeastern China, dated between 5.5–4.5 Myr ago.

The Beaver Pond sediments of Ellesmere Island record a fauna adapted to the most northerly terrestrial climates of the Early Pliocene. Those that have living descendants are all inhabitants of

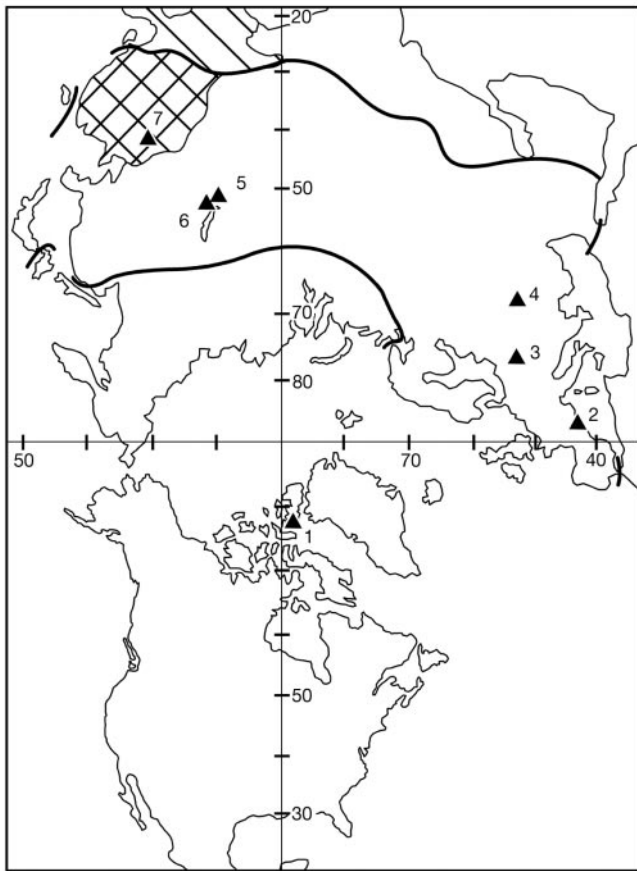


Figure 1 Numbered black triangles indicate sites producing *Arctomeles* species: (1) *A. sotnikovae* sp. nov., Ellesmere Island, Canada; (2) *A. genevauxi*, Montpellier, France; (3) *A. pliocaenicus* (genotype), Weze, Poland; (4) *A. ferus*, Odessa, Ukraine; (5) *A. suillus*, Shamar, Mongolia; (6) *A. suillus*, Udunga, Transbaykalia; (7) *A. suillus*, Yushe, China. Geographical ranges of living and fossil *Meles* are enclosed by a bold line; *Arctonyx* covered by slanting lines; overlap of genera by cross hatching. The units of values along the axes are degrees.

woodlands. One study¹² found that the moschid (their *Parablastomeryx* sp.) had dental microwear that indicated a “grazing diet”. The living *Moschus* feeds on grass as well as browse. The juvenile specimen of *Plesiohipparion* showed only wear associated with browsing, but this may not be indicative of adult feeding. Another study¹³ determined that the Beaver Pond *Hypolagus* was a cottontail ecomorph on the basis of the form of its astragalus. Thus the mammals indicate the presence of woodland with grassy understory providing food and shelter for elements of the fauna. The fossil flora from the site confirms this environmental range. Notably, the Beaver Pond evidence shows that *Dipoides*, representing a distinctly different clade than the living beaver *Castor*, appears to have shared the behaviour of cutting twigs and saplings. The presence of a mass of intertwined beaver-cut sticks up to 72 cm long and 5 cm wide together with cobbles in silty sand overlain by the peat suggests the possible core of a dam made by *Dipoides*.

Most of the Beaver Pond mammals have been found together in Early Pliocene deposits of the Yushe Basin, northeastern China, at 37°N latitude and 1,000 m elevation (Fig. 1, site 7). The accompanying pollen flora indicates a mixed conifer–hardwood forest containing spruce and pine, birch and alder as in the Beaver Pond, accompanied by *Ulmus*, and *Quercus* and such subtropical hardwoods as *Carya*, *Liquidambar* and *Magnolia*¹⁴. During the span from 4.4–4.0 Myr ago there was frequent alternation in the dominance of conifers and hardwoods in the Yushe Basin, but the climate remained warm and moist, comparable to that of the modern

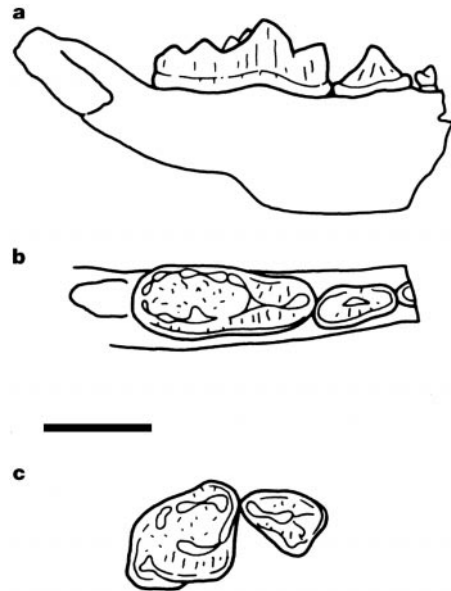


Figure 2 *Arctomeles sotnikovae* sp. nov. **a**, Lateral view of holotype ramal fragment with part of P₃, P₄–M₁. **b**, Occlusal view of same, alveolus for M₂ indicated. **c**, Occlusal view of paratype maxillary fragment (reversed) with P₄–M₁. Scale bar, 1 cm.

Qinling Mountains at 34°N latitude. Synthesis of global environments during the earlier Pliocene¹⁵ show greatly reduced continental ice volume, a seasonally ice-free Arctic Ocean and higher sea level. This flattening of present-day climatic gradients permitted the expansion of evergreen forest to the shore of the Arctic Ocean, and wider geographic ranges for Asian mammalian communities.

With the initiation of continental glaciation and permafrost at high latitudes at the close of the Gauss Chron (2.6 Myr ago), a profound faunal turnover occurred as taxa that characterize the Pleistocene Arctic became established¹⁶. Few of the clades represented in the Beaver Pond fauna are present in this new assemblage. □

Methods

Arctomeles Stach, 1951 (ref. 17)

Included species. *Arctomeles pliocaenicus* Stach, 1951 (genoholotype); *Meles genevauxi* Viret, 1939 (ref. 18); *Meles suillus* Teilhard and LeRoy, 1945 (ref. 19); *Parameles ferus* Roshchin, 1949 (ref. 20); *A. sotnikovae* sp. nov., this paper (Figs 1 and 2).

Revised generic diagnosis. Lower incisor crown triangular in cross-section; premolars small, low-crowned and separated by diastemata; M¹ lingual cingular shelf reduced, may be absent anteriorly; lower carnassial with short paraconid (foregoing characters are synapomorphies shared with *Arctonyx*); M₁ talonid longer and wider relative to length than in *Meles*; talonid basin filled with coarse crenulations.

Arctomeles sotnikovae sp. nov.

Etymology. For Marina Sotnikova, Geological Institute of Russia, Moscow, for her singular work on the fossil Carnivora of northeastern Asia.

Holotype. Canadian Museum of Nature (CMN) 51770 right ramus fragment with posterior half of P₃, P₄–M₁, M₂ alveolus (Fig. 2 a, b). Length and width of P₄, 7.7 and 3.8 mm; length of M₁, 15.6 mm; width of trigonid, 5.5 mm; width of talonid, 6.6 mm. **Paratype.** CMN 51771 left maxillary fragment with P₄–M₁ (Fig. 2c). Length and width of P₄, 7.3 and 5.4 mm; length and width of M₁, 10.0 and 9.2 mm.

Diagnosis. Smallest known species; P₄ large relative to M₁; P₄ with salient protocone; M¹ lingual cingular shelf extends to anterior end of tooth; M₁ entoconid closely associated with metaconid; hypoconulid small, no larger than more posterior talonid cusps.

Remarks. The new taxon is more primitive than other described species in the morphology of the P⁴ protocone, which is connate rather than shelf-like, and in having a continuous lingual cingulum on M¹. This morphology stands closer to that of Early Pleistocene *Meles* species such as *M. thoralis* of Europe than to later Pleistocene and living badgers, suggesting that cladogenesis giving rise to the living genera lay close to the Miocene/Pliocene boundary.

Received 24 March; accepted 4 July 2003; doi:10.1038/nature01892.

1. Fyles, J. G. High terrace sediments, probably of Neogene age, west-central Ellesmere island, Northwest Territories. *Curr. Res. Geol. Surv. Can.* 89-1D, 101–104 (1989).

2. Matthews, J. V. Jr & Oviden, L. E. Late Tertiary plant macrofossils from localities in Arctic/Subarctic North America: A review of the data. *Arctic* **43**, 364–392 (2000).
3. Elias, S. A. & Matthews, J. V. Jr Arctic North American seasonal temperatures from the latest Miocene to the early Pleistocene, based on mutual climatic range analysis of fossil beetle assemblages. *Can. J. Earth Sci.* **39**, 911–920 (2002).
4. Harington, C. R. *Life at a 3.5 Million Year Old Beaver Pond in the Canadian Arctic Islands and the Modern Scene* 11–13 (Meridian, Ottawa, 2001).
5. Hutchison, J. H. & Harington, C. R. A peculiar new fossil shrew (Lipotyphla, Soricidae) from the High Arctic of Canada. *Can. J. Earth Sci.* **39**, 439–443 (2002).
6. Zakrzewski, R. J. & Harington, C. R. Unusual Pliocene rodent from the Canadian Arctic Islands. *J. Vert. Paleol.* **21**, 116A–117A (2001).
7. Björk, P. R. The Carnivora of the Hagerman Local Fauna (Late Pliocene) of southwestern Idaho. *Am. Phil. Soc. Trans.* **60**, 1–54 (1970).
8. Lindsay, E. et al. Recognition of the Hemphillian/Blancan boundary in Nevada. *J. Vert. Paleol.* **22**, 429–442 (2002).
9. Petter, G. Origine, phylogénie et systématique des Blaireaux. *Mammalia* **35**, 567–597 (1971).
10. Hulbert, R. C. & Harington, C. R. An early Pliocene hipparionine horse from the Canadian Arctic. *Paleontology* **42**, 1017–1025 (1999).
11. Tedford, R. H. et al. Yushe Basin, China: Paleomagnetically calibrated mammalian biostratigraphic standard for the late Neogene of eastern Asia. *J. Vert. Paleol.* **11**, 519–526 (1991).
12. Dompierre, H. & Harington, C. R. Dental microwear analysis of an Early Pliocene deerlet and hipparionine from Strathcona Fjord, Ellesmere Island, Canada. *J. Vert. Paleol.* **17** (suppl. 3), 43A (1997).
13. Rybczynski, N. & Harington, C. R. Tarsal evidence for ecomorph reconstruction in fossil lagomorphs. *J. Vert. Paleol.* **17** (suppl. 3), 72A (1997).
14. Shi, N. The late Cenozoic stratigraphy, chronology, palynology and environmental development in the Yushe Basin, North China. *Striae* **36**, 1–90 (1994).
15. Dowsett, H. J. et al. *Middle Pliocene Paleoenvironmental Reconstruction: PRISM2* 99–535 (US Geol. Surv. Rep., Washington DC, 1999).
16. Sher, A. V. Olyorian land mammal age of northeastern Siberia. *Palaeontographica Italica* **74**, 77–112 (1986).
17. Stach, J. *Arctomeles pliocaenicus* new genus and species from Weze. *Acta Geol. Polonica* **2**, 129–157 (1951).
18. Viret, J. Monographie paléontologique de la faune de vertébrés des Sables de Montpellier, III, Carnivora, Fissipedia. *Trav. Lab. Geol. Fac. Sci. Lyon Mem.* **2**, 7–26 (1934).
19. Teilhard de Chardin, P. & Leroy, P. Les Mustélidés de Chine. *Inst. Geobiol. Pékin* **12**, 1–56 (1945).
20. Roshchin, A. D. A new genus of fossil mammal of the family Mustelidae from the Pliocene of Odessa. *Trudy Odeskogo Gosud. Univ.* **60**, 97–106 (1949).

Acknowledgements C.R.H. is grateful to J. G. Fyles, who discovered the Beaver Pond site in 1961 and the first vertebrate remains in 1988, for introducing him to the site in 1992, and to J. S. Tener and C. C. Kennedy for field assistance. The Canadian Museum of Nature and the Polar Continental Shelf Project provided support for field work. R.H.T. thanks E. Heck for graphics and A. Lora for manuscript preparation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.H.T. (tedford@amnh.org).

Volunteering leads to rock–paper–scissors dynamics in a public goods game

Dirk Semmann, Hans-Jürgen Krambeck & Manfred Milinski

Department of Evolutionary Ecology, Max Planck Institute of Limnology, 24306 Plön, Germany

Collective efforts are a trademark of both insect and human societies¹. They are achieved through relatedness in the former² and unknown mechanisms in the latter. The problem of achieving cooperation among non-kin has been described as the ‘tragedy of the commons’, prophesying the inescapable collapse of many human enterprises^{3,4}. In public goods experiments, initial cooperation usually drops quickly to almost zero⁵. It can be maintained by the opportunity to punish defectors⁶ or the need to maintain good reputation⁷. Both schemes require that defectors are identified. Theorists propose that a simple but effective mechanism operates under full anonymity. With optional participation in the public goods game, ‘loners’ (players

who do not join the group), defectors and cooperators will coexist through rock–paper–scissors dynamics^{8,9}. Here we show experimentally that volunteering generates these dynamics in public goods games and that manipulating initial conditions can produce each predicted direction. If, by manipulating displayed decisions, it is pretended that defectors have the highest frequency, loners soon become most frequent, as do cooperators after loners and defectors after cooperators. On average, cooperation is perpetuated at a substantial level.

Clean air to sustain the global climate and clean public toilets are examples of public resources that everybody is free to overuse. The social dilemma of public goods situations is that although a group of cooperators is always better off than a group of defectors, defectors exploit cooperators in groups. Since the late 1970s, economists, social scientists and evolutionary biologists have used the public goods game as a model to study the problem of maintaining cooperation in a group of unrelated individuals^{10–14}. For example, six players are asked to contribute money to a public pool; the money in the pool is, for example, multiplied by 3.6 and then equally distributed among the players irrespective of whether they contributed. The optimum outcome for the group is achieved if everybody cooperates; however, because each euro paid into the pool yields only a return of 60 cents for the contributor—that is, a net deficit of €0.40—no matter how the other players decide, the selfish decision is never to contribute to the pool. Studies have identified punishment^{6,15–17}, which is also combined with fairness¹⁸, and reputation through interaction with other social behaviour⁷ as mechanisms that can effectively maintain cooperation in public goods experiments.

In one model^{8,9}, a large population with three types of player, cooperators, defectors and loners, is considered. From time to time, sample groups of N players are randomly chosen and asked to participate in a single public goods game. Players either can refuse to participate, and will then receive a small fixed payoff, or can join the public goods game. In the latter case, they either defect or cooperate. Their strategies are specified beforehand and do not depend on the composition of the group. A continuing oscillation of the three strategies is predicted because each strategy, when most frequent, can be beaten by one of the others. Defectors can exploit a large group of cooperators, whereas loners have the highest profit when defectors are frequent. When loners are most frequent the public group size is reduced, which invites cooperation because the game is no longer a dilemma in small groups^{19–21}. For example, if the group consists of only three players, each euro paid into the public pool yields a return of €1.20 for the contributor, that is, a net gain of €0.20.

It is not just the fact that volunteering is possible that induces cooperation, but rather that volunteering reduces public goods groups to small sizes for which the individual cost-to-benefit ratio becomes more favourable. In addition, even though defectors are still better off than are cooperators in each group, cooperators do better when averaged over small groups according to Simpson’s paradox²². For example, a group of three players can consist of either three cooperators, two cooperators and one defector, one cooperator and two defectors, or three defectors; cooperators receive on average a net gain of €1.8, defectors only €0.8. Circumstantial evidence for the ‘small group advantage’ is potentially provided by fish that leave their shoal and take a risk to inspect a predator from a short distance: very often minnows, *Phoxinus phoxinus*, inspect a pike, *Esox lucius*, in small groups²³. Thus, after loners, cooperators will be most frequent for a while before defectors will take over again^{8,9}. Hence, volunteering relaxes the social dilemma: instead of defectors winning the world, coexistence among cooperators, defectors and loners is expected²⁴.

We tested these predictions with 280 first-semester biology students in 20 groups of 14 students that played the optional public goods game for 57 consecutive rounds. The students observed the

introduction and the complete game on a public screen. They were told that they had a starting account of €10 and would make their decisions anonymously. In each round, six players were randomly selected from the 'population' of 14 players to decide first whether to join the public goods group and, thereafter, if they chose to join whether to contribute to the public pool. In the first seven rounds, we manipulated the displayed decisions in such a way that defector, cooperator or loner was pretended to be the most frequent strategy of the population. This manipulation was necessary to test the three possible predictions of the model experimentally. Without this manipulation our results would be only descriptive. In the eighth round, we expected that being a loner (after staged defector), defector (after staged cooperator) or cooperator (after staged loner) would be the most frequent strategy according to the players' real decisions. Thereafter, the game proceeded with unmanipulated display to test whether oscillations of the three strategies would occur and, if so, whether they would occur predominantly in the predicted sequence over 50 rounds.

After the manipulated start in the first seven rounds, we found that the predicted strategy was the most frequent strategy in round eight after all three starting schemes ($P < 0.004$, $n = 20$ groups, sign test, two-tailed; Fig. 1a–c). We use each group of 14 players as a statistical unit. In the following 50 rounds, we determined for each group of 14 players the number of cases where two conditions were met: one strategy was most frequent and one of the other strategies was most frequent in the following round. We compared all cases where the predicted strategy became most frequent with all cases where an unpredicted strategy became most frequent. In this way, we identified switches of the most frequent strategy between rounds and checked whether their direction was as predicted. The predicted

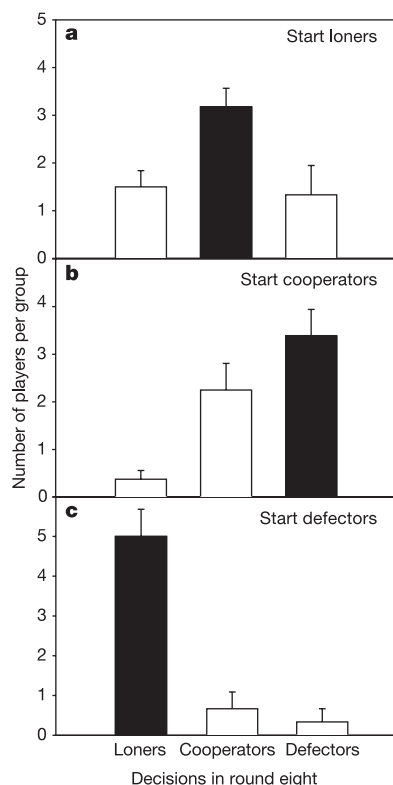


Figure 1 Decisions in round eight after the staged standstill with one single strategy that was most frequent in the first seven rounds. The bars of the strategy predicted to be most frequent are shown in black. The average frequency of chosen strategies per group of 14 players is shown (mean $\pm$ s.e.m.). **a**, Start loners: $n = 6$ groups of 14 players with a simulated prevalence of loners in rounds 1–7. **b**, Start cooperators: $n = 8$ groups of 14 players with a simulated prevalence of cooperators in rounds 1–7. **c**, Start defectors: $n = 6$ groups of 14 players with a simulated prevalence of defectors in rounds 1–7.

strategy became most frequent significantly more often than did the alternative strategy ($P < 0.001$, $n = 20$; paired t -test, $t = 6.588$, two-tailed; Fig. 2).

Although the above analysis provided a formal proof of the predicted oscillations, we made an example of these oscillations visible in Fig. 3. We synchronized the 20 groups in the 50 'not-manipulated' rounds by selecting similar starting points in each group, because the cycles were not expected to have the same duration in each group. For example, if we were to select the round from each group that had the highest proportion of loners, we would expect that cooperators would be most frequent next, followed by defectors in all 20 groups. The same procedure was used to find such starting points for cooperators and for defectors.

From several maxima of a strategy, we defined the first as the starting point. Thereafter, we averaged all 20 starting point rounds and each of the following 9 rounds over all groups, for loners (Fig. 3a), cooperators (Fig. 3b) and defectors (Fig. 3c) as starting points. The oscillations can be observed in all three cases, although the groups became increasingly asynchronous during the ten rounds. As the model predicts, after loners have the highest frequency, cooperators become most frequent, thereafter defectors, and then loners again (Fig. 3a). After a prevalence of cooperators, defectors become most frequent, followed by loners, and then cooperators again (Fig. 3b). Figure 3c shows that the prevalence of defectors is followed, as predicted, by an increase of loners that is closely trailed by increasing numbers of cooperators, followed again by defectors, and thereafter by loners.

The consequences of the oscillation of the strategies should be an always-recurring rise of each of the three strategies and thus a fairly cooperative outcome of the game after initial perturbations. In the last 30 rounds (21–50), the frequencies of the three strategies seemed on average rather stable (rounds 21–35: $32.22 \pm 1.0\%$ loners, $30.11 \pm 0.9\%$ cooperators, and $37.67 \pm 1.0\%$ defectors; rounds 36–50: $32.39 \pm 1.4\%$ loners, $29.06 \pm 1.3\%$ cooperators, and $38.56 \pm 1.3\%$ defectors). According to the model, we would expect that there would be at least 42% loners and that 58% would choose to join the public goods group⁹. Only $33 \pm 2.5\%$ (mean $\pm$ s.e.m.) chose the loner option, which is significantly less than expected ($P = 0.003$, $n = 20$; Wilcoxon signed rank test, $Z = 2.95$, two-tailed). Of the players joining the public goods group, 38% would be expected to cooperate and 62% to defect, respectively⁹. We found, as expected, more defectors ($56.51 \pm 1.7\%$) than cooperators ($43.48 \pm 1.7\%$, $P = 0.004$, $n = 20$; Wilcoxon signed rank test, $Z = 2.91$, two-tailed). Although these numbers are close to the expected ones, the percentage of cooperators was significantly higher than predicted ($P = 0.011$, $n = 20$; Wilcoxon signed rank test, $Z = 2.54$, two-tailed).

In the long run (averaging over many cycles), the net payoff of both defectors and cooperators should be same as the payoff of loners—that is, €1.25. We found that defectors earned slightly but significantly more than expected, $1.46 \pm €0.04$ ($P < 0.001$, $n = 20$;

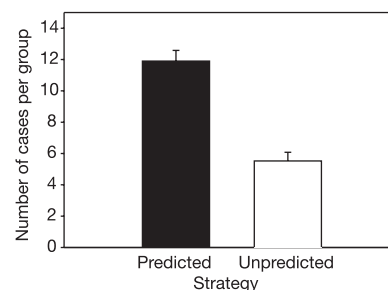


Figure 2 The predicted prevalence switch occurred more frequently than the unpredicted prevalence switch during the 50 rounds that followed the seven manipulated start rounds. Columns show mean $\pm$ s.e.m. per group of 14 players.

Wilcoxon signed rank test, $Z = 3.36$, two-tailed). Cooperators had a payoff that did not significantly differ from the expected one, $1.32 \pm €0.09$ ($P = 0.43$, $n = 20$; Wilcoxon signed rank test, $Z = 0.78$, two-tailed). Defectors probably profited because they were less frequent than expected at the equilibrium.

We found that volunteering (the option to choose between joining the public goods group and taking the loner strategy) indeed protected cooperation in the public goods game by inducing small group sizes. On average, there was a rather stable frequency of cooperators that was higher than what is usually found in public goods games after several rounds^{5,17}. As predicted by the model^{8,9}, the dynamics of the games showed oscillations of the rock–paper–scissors succession of cooperators, defectors and loners, even though our players were less averse to risk than expected: only about a third chose the loner option.

Volunteering is a mechanism that potentially sustains cooperation in various species. Like some large predatory animals, ancestral humans also acted as groups when hunting large prey such as mammoths, but went out solitarily for small prey such as antelopes²⁵. Thus, volunteering was possible and might have supported cooperation in addition to potential relatedness by reducing the public (hunting) group size. Obviously, we are not free to decide whether we stop sharing the global climate with others, but there are many other human social dilemmas in which volunteering is possible.

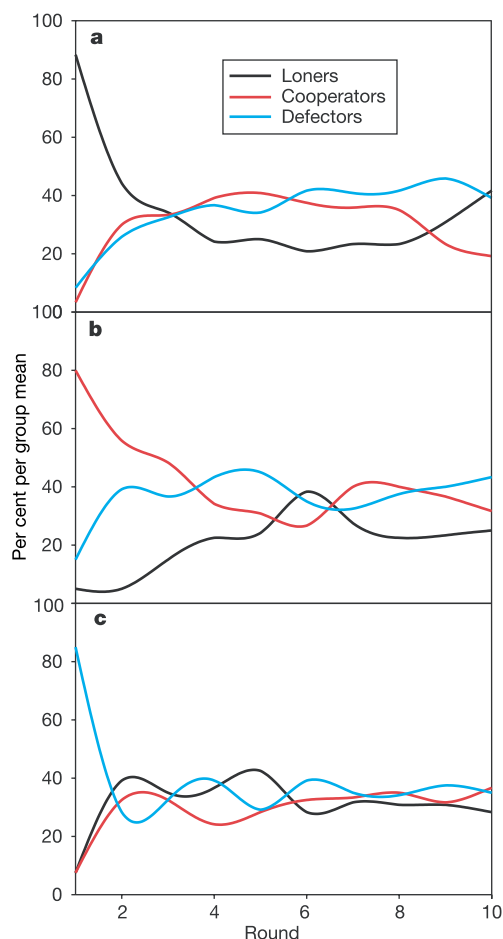


Figure 3 Average frequencies of the three strategies over a period of ten rounds after synchronizing the 20 groups as follows. The starting round of the ten rounds was defined for each group as the round when one of the strategies reached its maximum proportion throughout the game (rounds 8–57) for the first time. All 20 starting rounds and each of the following 9 rounds were averaged over all groups for loners (a), cooperators (b) and defectors (c). The sequence follows the predicted most frequent strategies according to rock–paper–scissors dynamics: that is, loners, cooperators, defectors, loners, and so on.

Volunteering does not produce overwhelming cooperation, but it might help to avoid the fate of mutual defection in many human collective enterprises and thus might pave the way for other mechanisms of cooperation to take over. For example, direct²⁶ or indirect reciprocity^{27–30} may be catalysed when the population happens to be in a cooperator period of the rock–paper–scissors dynamic and anonymity is relaxed after repeated interactions. Loners, although unsocial by definition, help cooperators to become most frequent and thus to escape the social dilemma. □

Methods

Subjects

A total of 280 human subjects of the universities of Bonn, Hamburg and Kiel played a public goods game with optional participation that lasted for 57 rounds. The students were completely anonymous, sat between partitions, saw the introduction to the game including one example round and the complete game on a large screen. They did not know the total number of rounds. They interacted by means of a computer program with silent 'yes' and 'no' switches.

Basis of the public goods game

For each round the computer program randomly selected 6 of the 14 students. Each student had played almost the same number of rounds at the end of the game. Because the expected cycles are predicted to become smoother with increasing population size⁹, we mimicked a larger population. The students were told that there was a pool of additional players in the form of strategies recorded from earlier sessions and that the program would sometimes choose 'players' from this pool.

A light at each person's desk signalled who was to decide. Each of the six players had to decide first whether to play the loner strategy, thereby obtaining a fixed payoff (€1.25), or to join the public goods group with a second decision to make. The minimum public group size was two players. If only one player decided to play in the public goods group, he/she knew that he/she would automatically also become a loner. If the public goods group size was either two or larger, the players that had chosen to play in the public goods group had to decide whether they would contribute €1.25 or nothing to the public pool. At this point, they did not know how many subjects had decided to play in the public goods group. After all players of the public goods group made their final decision, the content of the pool was multiplied by 3.6 and divided evenly among the players that had joined the public goods group irrespective of their actual contribution. With an interest rate of 3.6, the model system has a fixed point, which refers to substantial proportions of cooperators, defectors and loners. The dynamic then predicts periodic cycles of all three strategies around these proportions; this requires an interest rate larger than 2.

Only at this point were the decisions of all players displayed simultaneously on the screen that all 14 subjects could see: that is, the numbers of loners and public good group players, their payoffs and their eventual costs (for example, one player was a loner and obtained €1.25 without cost, five had chosen to join the public goods group, of which three were defectors who received a payoff of €1.80 from the pool without costs and two were cooperators who also received €1.80 from the pool, but they had costs of €1.25 each). It never happened that one subject had to play loner because he/she had no money left.

Rounds of the public goods game

In the first seven rounds, the display was manipulated such that the players were led to believe that they were in a group that played a high percentage of only one strategy. In six groups loners appeared to be most frequent; there were eight groups with cooperators and six with defectors as the apparent most frequent strategy. The players could make decisions, however, which were not displayed. Instead, six predetermined decisions with corresponding payoffs and eventual costs were shown. Each of the three possible real decisions of a player (that is, loner, cooperator and defector) was included at least once to ensure that each player would find his actual decision on the screen and nobody would doubt that the displayed decisions were real. The maximum number of defectors or cooperators displayed on the screen was four per round; for example, four defectors, one cooperator, one loner. In the case of loners prevailing in the first seven rounds, it was possible to show up to 100% of loners in one round, because each player who decided to join the public goods group would believe that he/she was the only one with this decision and thus became a loner. To have some variation, we chose a percentage of loners that was slightly lower in the seven rounds. On average, there were 79% of loners in the staged loner groups, 61% of cooperators in the staged cooperator groups, and 64% of defectors in the staged defector groups. Starting with round eight, there was no manipulation of the display for 50 consecutive rounds. The students did not know the total number of rounds to be played.

Received 30 June; accepted 6 August 2003; doi:10.1038/nature01986.

1. Trivers, R. *Social Evolution* (Benjamin Cummings, Menlo Park, California, 1985).
2. Hamilton, W. D. Genetical evolution of social behaviour. *J. Theor. Biol.* **7**, 1–52 (1964).
3. Hardin, G. Tragedy of commons. *Science* **162**, 1243–1248 (1968).
4. Hardin, G. Extensions of 'the tragedy of the commons'. *Science* **280**, 682–683 (1998).
5. Ledyard, J. O. in *Handbook of Experimental Economics* (eds Kagel, J. H. & Roth, A. E.) 111–194 (Princeton Univ. Press, Princeton, New Jersey, 1995).
6. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
7. Milinski, M., Semmann, D. & Krambeck, H. J. Reputation helps solve the 'tragedy of the commons'. *Nature* **415**, 424–426 (2002).
8. Hauert, C., De Monte, S., Hofbauer, J. & Sigmund, K. Volunteering as red queen mechanism for cooperation in public goods games. *Science* **296**, 1129–1132 (2002).

9. Hauert, C., De Monte, S., Hofbauer, J. & Sigmund, K. Replicator dynamics for optional public good games. *J. Theor. Biol.* **218**, 187–194 (2002).
10. Sugden, R. *The Economics of Rights, Co-operation and Welfare* (Blackwell, Oxford, UK, 1986).
11. Ostrom, E. *Governing the Commons* (Cambridge Univ. Press, Cambridge, 1999).
12. Gintis, H. *Game Theory Evolving* (Princeton Univ. Press, Princeton, New Jersey, 2000).
13. Berkes, F., Feeny, D., McCay, B. J. & Acheson, J. M. The benefits of the commons. *Nature* **340**, 91–93 (1989).
14. Colman, A. M. *Game Theory and Its Applications in the Social and Biological Sciences* (Butterworth-Heinemann, Oxford, UK, 1995).
15. Boyd, R. & Richerson, P. J. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol. Sociobiol.* **13**, 171–195 (1992).
16. Gintis, H. Strong reciprocity and human sociality. *J. Theor. Biol.* **206**, 169–179 (2000).
17. Fehr, E. & Gächter, S. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **90**, 980–994 (2000).
18. Fehr, E. & Rockenbach, B. Detrimental effects of sanctions on human altruism. *Nature* **422**, 137–140 (2003).
19. Schelling, T. C. Hocky helmets, concealed weapons, and daylight saving—study of binary choices with externalities. *J. Confl. Resolut.* **17**, 381–428 (1973).
20. Dawes, R. M. Social dilemmas. *Annu. Rev. Psychol.* **31**, 169–193 (1980).
21. Boyd, R. & Richerson, P. J. The evolution of reciprocity in sizable groups. *J. Theor. Biol.* **132**, 337–356 (1988).
22. Sober, E. & Wilson, D. S. *Unto Other: The Evolution and Psychology of Unselfish Behavior* (Harvard Univ. Press, Cambridge, Massachusetts, 1999).
23. Magurran, A. E. & Pitcher, T. J. Provenance, shoal size and the sociobiology of predator-evasion behaviour in minnow shoals. *Proc. R. Soc. Lond. B* **229**, 439–465 (1987).
24. Michor, F. & Nowak, M. A. Evolution - the good, the bad and the lonely. *Nature* **419**, 677–679 (2002).
25. Ridley, M. *The Origins of Virtue* (Penguin, London, 1996).
26. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
27. Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577 (1998).
28. Wedekind, C. & Milinski, M. Cooperation through image scoring in humans. *Science* **288**, 850–852 (2000).
29. Bolton, G. E., Katok, E. & Ockenfels, A. What's in a reputation? Indirect reciprocity in an image scoring game. Working paper. (Penn State University, 2001).
30. Seinen, I. & Schram, A. *Social status and group norms: indirect reciprocity in a helping experiment. Discussion paper TI2001-003/1* (Tinbergen Institute, Amsterdam, 2001).

Acknowledgements We thank students at the universities of Bonn, Kiel and Hamburg for participation; T. Bakker, H. Brendelberger, E. Heinz, K.-P. Sauer and M. Zbinden for support; C. Hauert for calculating parameters; and the Max-Planck-Institute of Meteorology for hospitality.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.M. (milinski@mpil-ploen.mpg.de).

Cytosolic pH regulates root water transport during anoxic stress through gating of aquaporins

Colette Tournaire-Roux¹, Moira Sutka¹, Hélène Javot¹, Elisabeth Gout², Patricia Gerbeau^{1*}, Doan-Trung Luu¹, Richard Bligny² & Christophe Maurel¹

¹Biochimie et Physiologie Moléculaire des Plantes, Centre National de la Recherche Scientifique (Unité Mixte de Recherche 5004), Institut National de la Recherche Agronomique, Université Montpellier 2 et Ecole Nationale d'Agronomie, Place Viala, F-34060 Montpellier cedex 1, France
²Physiologie Cellulaire Végétale, Commissariat à l'Energie Atomique, Rue des Martyrs, F-38054 Grenoble cedex 9, France

* Present address: Phytobiologie Cellulaire, Université de Bourgogne, BP 47 870, F-21078 Dijon cedex, France

Flooding of soils results in acute oxygen deprivation (anoxia) of plant roots during winter in temperate latitudes, or after irrigation¹, and is a major problem for agriculture. One early response of plants to anoxia and other environmental stresses is down-regulation of water uptake due to inhibition of the water permeability (hydraulic conductivity) of roots (L_p)^{2–5}. Root water uptake is mediated largely by water channel proteins

(aquaporins) of the plasma membrane intrinsic protein (PIP) subgroup^{6–8}. These aquaporins may mediate stress-induced inhibition of L_p ^{2,4,9} but the mechanisms involved are unknown. Here we delineate the whole-root and cell bases for inhibition of water uptake by anoxia and link them to cytosol acidosis. We also uncover a molecular mechanism for aquaporin gating by cytosolic pH. Because it is conserved in all PIPs, this mechanism provides a basis for explaining the inhibition of L_p by anoxia and possibly other stresses. More generally, our work opens new routes to explore pH-dependent cell signalling processes leading to regulation of water transport in plant tissues or in animal epithelia¹⁰.

The molecular bases of aquaporin gating in plants and animals by phosphorylation^{11–13} or other mechanisms^{14,15} remain elusive. Certain mammalian aquaporins are regulated by external pH^{14–16} when expressed in *Xenopus* oocytes. In plants, the water channel activity of purified plasma membrane vesicles can be blocked by protons¹⁷. We investigated the relevance of this process for regulation of water uptake by roots in *Arabidopsis*.

Roots detached from plants grown in hydroponics were inserted into a pressure chamber and bathed in a well-aerated standard root bathing solution (RBS) at pH 6.0. Applied pressure (P) induced a flow (J_v) of exuded sap, with a linear J_v to P relationship for P up to 0.5 MPa. The slope reported to root dry weight indicates a mean L_p of $61.5 \pm 6.7 \text{ ml g}^{-1} \text{ h}^{-1} \text{ MPa}^{-1}$ ($\pm \text{s.e.m.}$; $n = 4$) (Fig. 1a). Oxygen deprivation was induced in the same roots by N_2 bubbling for 30 min in RBS and resulted in a 42% reduction in L_p to $35.6 \pm 7.4 \text{ ml g}^{-1} \text{ h}^{-1} \text{ MPa}^{-1}$ ($\pm \text{s.e.m.}$) (Fig. 1a). Because the J_v curves cross the P axis at similar values close to the origin, measurements of J_v at a high (>0.3 MPa) constant P can be used to monitor relative changes in L_p . For instance, the anoxic treatment above induced a 46% reduction in J_v at 0.35 MPa. A similar reduction in J_v by $49.2 \pm 3.7\%$ ($\pm \text{s.e.m.}$, $n = 5$) was observed on another set of plants. Altogether, these results extend to *Arabidopsis* observations previously made in crop species^{3,5}. A fall in cytosolic pH is, besides fluctuations in cytosolic Ca^{2+} , one early cellular response that typically accompanies anoxia^{18–20}. This response was investigated in *Arabidopsis* roots using *in vivo* proton-decoupled ³¹P-nuclear magnetic resonance (NMR) spectroscopy. Spectra (see Supplementary Information 1) revealed three major peaks corresponding to phosphorylcholine (P-Cho), at 3.47 parts per million (p.p.m.), and to cytosolic and vacuolar pools of inorganic phosphate (Pi), at 2.45 and 0.55 p.p.m.—pH values of 7.7 and 5.9, respectively. Within the 10 min after the onset of oxygen deprivation, the peaks of cytosolic Pi and P-Cho shifted upfield to 1.94 p.p.m. and 3.38 p.p.m., respectively, indicating a decrease in cytosolic pH to 7.20–7.25 ($n = 3$).

Inhibitors of cytochrome pathway respiration can be used to mimic oxygen deprivation^{4,5}. We found that sodium azide (1 mM NaN_3) and potassium cyanide (0.5 mM KCN) induced a marked inhibition of J_v by $87 \pm 1\%$ (NaN_3 ; $n = 6$) and $81 \pm 1\%$ (KCN; $n = 7$), with half times ($t_{1/2}$) of 2.6 ± 0.2 min and 4.6 ± 0.1 min, respectively (Fig. 1b). Wash out of NaN_3 and KCN induced a significant reversal of J_v inhibition to $92 \pm 3\%$ ($n = 3$) and $76 \pm 3\%$ ($n = 4$) of initial values, respectively, over a similarly short period of time (NaN_3 , $t_{1/2} = 4.6 \pm 0.6$ min; KCN, $t_{1/2} = 3.2 \pm 0.7$ min; Fig. 1b). In parallel NMR experiments the chemical shift of P-Cho indicated that exposure of roots to 1 mM NaN_3 or 0.5 mM KCN resulted in a rapid drop in cytosolic pH to a steady-state value of 6.9–7.0 (NaN_3) or 7.0–7.25 (KCN; $n = 3$) after 10–20 min (Fig. 1c; see also Supplementary Information 1). Wash out of the inhibitors resulted in a complete recovery of initial cytosolic pH values (pH 7.7) in less than 5 min (Fig. 1c).

Loading of root cells with 20 mM propionic acid/potassium propionate (KProp), by substituting at constant pH (pH 6.0) the weak acid for 20 mM KCl in a standard RBS, induced a marked drop in cytosolic pH down to 6.5–6.6 ($n = 3$), with a secondary equilibration to steady-state values in the range of pH 6.7–6.8 (Fig. 2b). A

rapid recovery to initial cytosolic pH values was observed on return to standard RBS (Fig. 2b). Parallel pressure chamber measurements indicated that a similar acid load treatment induced a rapid ($t_{1/2} = 3.7 \pm 0.3$ min; $n = 6$) and significant reduction ($71 \pm 3\%$) in J_v (Fig. 2a). On subsequent transfer into control RBS, a significant recovery from J_v inhibition ($70 \pm 6\%$; $n = 4$) occurred rapidly ($t_{1/2} = 1.8 \pm 0.2$ min), before J_v slowly decreased again and stabilized to a slightly reduced, steady-state value (Fig. 2a). Altogether, kinetic measurements of J_v and cytosolic pH under acid load and respiration blockade conditions establish a strong corre-

lation between cytosolic pH and root water transport capacity. The partially irreversible alterations of L_p , observed after acid load may result from the large amplitude changes in cytosolic and vacuolar pH values and associated metabolic disturbances. Also, root cells that have the most abundant cytosolic Pi pools and thus contribute to most of the NMR signal may not represent the critical barriers for water transport regulation. This may explain further slight discrepancies between kinetic changes in cytosolic pH and J_v . Furthermore, changing the pH between 5.5 and 8.0 of a RBS had no effect on cytosolic pH and did not noticeably alter L_p , indicating that, by

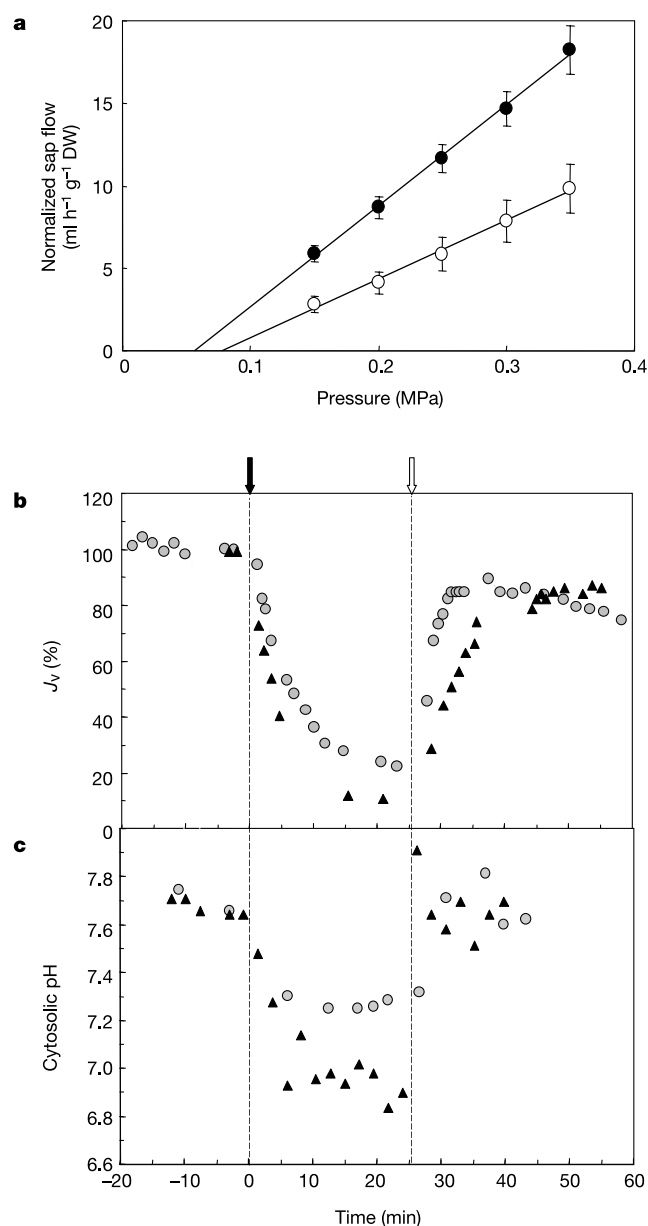


Figure 1 Effects of anoxia and respiratory inhibitors on *Arabidopsis* root water transport and cytosolic pH. **a**, Pressure chamber measurements ($\pm$ s.e.m.) on individual excised root systems ($n = 4$) bathed in a standard RBS before (normoxic, filled circles) and after (anoxic, open circles) a 30-min bubbling treatment with N_2 . J_v was expressed per unit dry weight and plotted as a function of applied pressure. The slope of the regression lines corresponds to the root hydraulic conductivity (L_p). **b**, Representative kinetic changes of J_v (% of initial value) induced on two individual excised root systems by exposure to 1 mM NaN_3 (filled triangles) or 0.5 mM KCN (filled circles). Inhibitors were added at $t = 0$ (filled arrow) and removed at $t = 25$ min (open arrow). The applied pressure was kept constant at 0.3 MPa. **c**, Kinetic changes in cytosolic pH in response to exposure to respiratory inhibitors. Same conventions as in **b**.

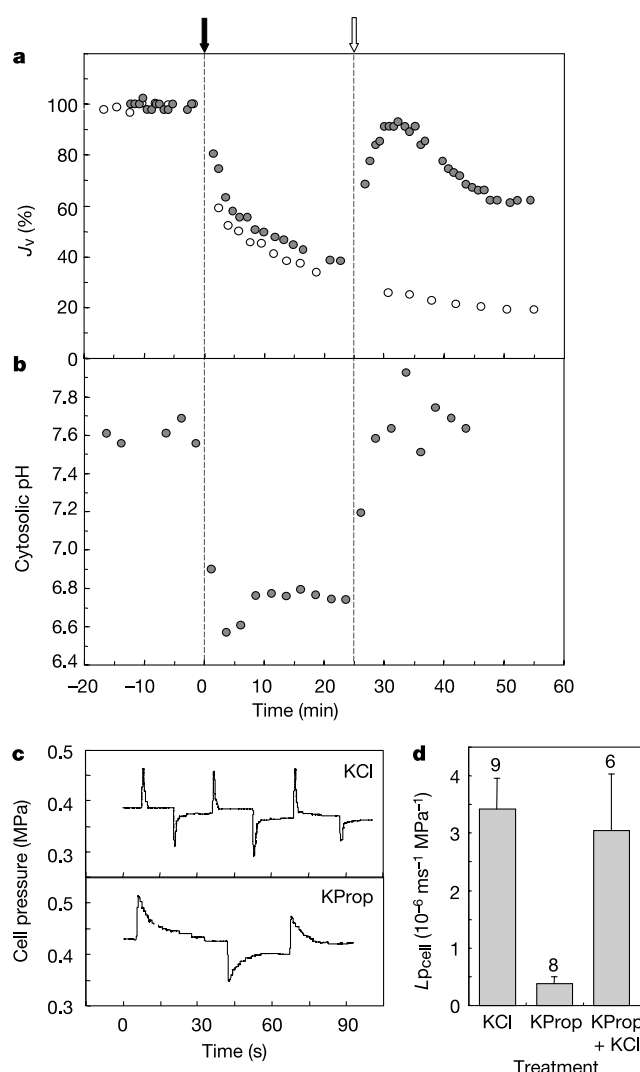


Figure 2 Effects of an acid load on cytosolic pH and water transport in whole roots and cortex cells. **a**, Kinetic changes in J_v induced in response to an acid load with 20 mM propionic acid at pH 6.0. Acid load was applied at $t = 0$ (filled arrow) and maintained (open circles) or reversed (filled circles) at $t = 25$ min (open arrow). **b**, Acid-load-induced changes in cytosolic pH. Same treatment and conventions as in **a**. Cytosolic pH was deduced from the chemical shift of P-Cho determined in 2.5-min acquisition spectra. **c**, Typical time course of root cortex cell pressure relaxation in control (KCl) and acid-loaded (KProp) roots. After measurements of a stationary turgor pressure, exo-osmotic (outward) and endo-osmotic (inward) water movements were triggered by quickly changing the cell pressure using the cell probe, and then hydrostatic relaxations at constant probe volume (meniscus immobilized manually) were recorded. For the individual cells shown, mean water relation parameters were: KCl, $t_{1/2} = 0.50$ s, $L_{p\text{cell}} = 4.16 \times 10^{-6} \text{ ms}^{-1} \text{ MPa}^{-1}$; KProp, $t_{1/2} = 3.98$ s, $L_{p\text{cell}} = 0.44 \times 10^{-6} \text{ ms}^{-1} \text{ MPa}^{-1}$. **d**, Mean $L_{p\text{cell}}$ values for the indicated number of cells from root segments that were pre-treated for 12.5–22.5 min in a KCl or a KProp solution or successively treated for 20 min in KProp and 15–20 min in a KCl solution (KProp + KCl).

contrast to cytosolic pH, extracellular pH does not regulate water transport in *Arabidopsis* roots (data not shown).

Root water transport is mediated through the parallel contribution of cell wall (apoplastic) and cell-to-cell paths^{21,22}. The latter includes transport across cellular membranes and/or plasmodesmata and is a likely target for reversible regulation by cytosolic pH. To investigate this, the hydraulic conductivity of root cortex cells ($L_{p\text{cell}}$) was measured by the cell pressure probe technique^{6,23} in conditions similar to those used for measuring J_v and cytosolic pH. Cells of roots bathed in a control saline solution with 20 mM KCl, pH 6.0, displayed typically short hydrostatic equilibration times ($t_{1/2} = 0.66 \pm 0.05$ s; $n = 9$; Fig. 2c) indicative of a high hydraulic conductivity ($L_{p\text{cell}} = 3.43 \pm 0.52 \times 10^{-6} \text{ m s}^{-1} \text{ MPa}^{-1}$; Fig. 2d). In contrast, acid load for 12.5–22.5 min with 20 mM KProp at pH 6.0 resulted in a marked decrease (89%) in $L_{p\text{cell}}$ (Fig. 2d). This decrease was reflected by a greater than tenfold increase in $t_{1/2}$ ($t_{1/2} = 7.29 \pm 1.24$ s; $n = 8$; Fig. 2c) without significant alteration in stationary turgor pressure nor in volumetric elastic modulus (data not shown). Cells from root segments that had been sequentially acid loaded and returned in a control saline solution had a mean $L_{p\text{cell}}$ similar to that of control untreated cells (Fig. 2d). Although cell analysis was only accessible in the cortex⁶, the amplitude of acid-load-induced inhibition of $L_{p\text{cell}}$ suggests that blockade of the trans-cellular path can account for most effects observed on whole-root water transport (see Supplementary Information 2).

Because they contribute significantly to $L_{p\text{cell}}$ in roots^{6,8}, PIP aquaporins may well account for the cytosolic pH-dependent component. The direct effects of extra- and intracellular acidification on the activity of specific PIP isoforms were investigated using heterologous expression in *Xenopus* oocytes. Measurements with a pH-sensitive electrode showed that oocytes maintain physiological cytosolic pH values in the 7.1–7.3 range when exposed to bathing solutions with 50 mM NaCl at pH 7.5 (NaCl 7.5) or pH 6.0 (NaCl 6.0) or with 50 mM sodium acetate at pH 7.5 (NaAc 7.5) (Fig. 3a). In contrast, exposure to 50 mM NaAc at pH 6.0 (NaAc 6.0) induced a

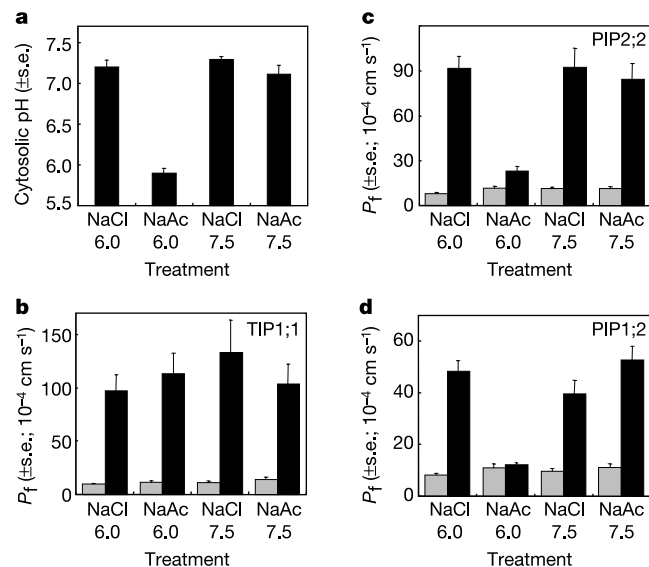


Figure 3 Effects of a stepwise acid load treatment on cytosolic pH and aquaporin water transport activity in *Xenopus* oocytes. **a**, Native oocytes ($n = 7$ – 8) were pre-treated for 10 min in solutions containing 50 mM NaCl or NaAc at pH 6.0 or pH 7.5, and cytosolic pH was measured in the same solution using a pH-sensitive electrode. **b**–**d**, After pre-treatments as in **a**, P_f of control uninjected (grey bars; $n \geq 11$) or cRNA-injected (black bars; $n \geq 19$; **b**, TIP1;1; **c**, PIP2;2; **d**, PIP1;2) oocytes was measured by a swelling assay in the same solutions but with a fivefold dilution in distilled water. Each panel represents pooled data from three independent batches of oocytes.

marked cytosolic acidification in less than 10 min (cytosolic pH = 5.90 ± 0.06 ; $n = 8$; Fig. 3a). In parallel osmotic swelling assays, we observed that the osmotic water permeability (P_f) conferred to the oocyte membrane by expression of *Arabidopsis* vacuolar membrane aquaporin TIP1;1 (ref. 24) was insensitive to a 10-min pre-treatment of oocytes by NaCl or NaAc at either pH 6.0 or 7.5 (Fig. 3b). In contrast, the activity of all PIP2 homologues investigated (PIP2;1, PIP2;2, PIP2;3) was markedly ($\geq 85\%$) and specifically inhibited by a NaAc 6.0 pre-treatment (Fig. 3c). Among several PIP1 isoforms investigated (PIP1;1, PIP1;2, PIP1;3, PIP1;4), only PIP1;2 displayed a significant water transport activity in oocytes, which again was selectively blocked in NaAc 6.0 (Fig. 3d).

The finding that most, if not all, PIPs are specifically blocked by intracellular acidification implies that PIPs have a conserved structural basis for cytosolic pH sensing. Incubation of oocytes expressing PIP2;2 in NaAc solutions with pH between 6.0 and 7.5 ($6.0 < \text{cytosolic pH} < 6.9$) indicated that the apparent pK of proton-dependent PIP inhibition was in the range of pH 6–6.9 (Fig. 4a). This and the sidedness of pH effects suggest that proto-

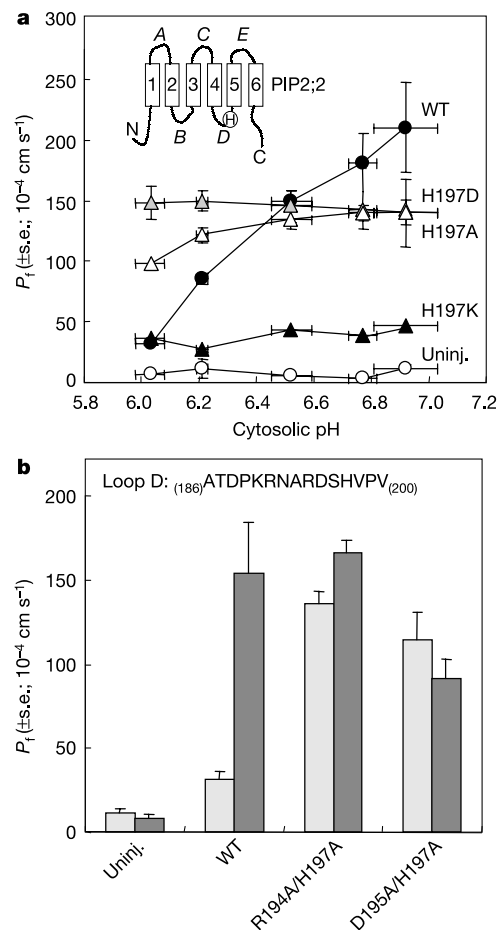


Figure 4 Sensitivity of wild-type (WT) and mutant PIP2;2 to cytosolic acidification.

a, Control uninjected oocytes (open circles) or oocytes expressing wild-type PIP2;2 (filled circles) or H197D (grey triangles), H197A (white triangles) or H197K (black triangles) mutants were pre-treated for 10 min in NaAc solutions with pH between 6.0 and 7.5. P_f values were measured and are reported as a function of cytosolic pH measured in control oocytes pre-treated in the same conditions. Error bars not shown fall into symbols ($n = 4$ – 6). The inset shows a schematic view of PIP2;2 topology and the position of H197 (H) in loop D. **b**, Control uninjected oocytes ($n = 9$ – 10) or oocytes injected with cRNAs ($n \geq 6$) encoding wild-type PIP2;2 or double PIP2;2 mutants R194A/H197A or D195A/H197A were pre-treated in a NaAc 6.0 (light grey bars; cytosolic pH of 6.0) or a NaAc 7.5 (dark grey bars; cytosolic pH of 6.9) solution and their P_f value was measured. The inset shows the sequence of loop D in PIP2;2.

nation of one or more cytoplasmically exposed His residues is involved. Sequence comparison indicated that all 13 *Arabidopsis* PIPs share a highly conserved pattern of five His residues (Supplementary Information 3). Although all PIP1 and certain PIP2 homologues have a few additional His residues at varying positions, PIP2;2 is one PIP isoform that exhibits the minimal five His pattern. Three of these face the cytosol with only two being specific for PIPs: His 197 and His 264, which are localized in intracellular loop D (Fig. 4a, inset) and at the end of the sixth membrane-spanning domain, respectively. In experiments where oocyte cytosolic pH was adjusted to values between 6.0 and 6.9, a PIP2;2 mutant with substitution of His 197 by an Ala residue (H197A) exhibited a much less pronounced blockade by cytosolic acidification (30%) than wild-type PIP2;2 (85%) (Fig. 4a). A H197D mutant carrying a negatively charged Asp at the same position exhibited a constitutively high, pH-insensitive activity. Furthermore, substitution of His 197 by a Lys residue (H197K) conferred a totally pH-insensitive but low apparent activity, in agreement with the assumption that the introduced positive charge may mimic the protonated state of His 197 (Fig. 4a). Altogether, the effects of point mutations and/or charge substitution at position 197 establish His 197 of PIP2;2 as a major site for cytosolic pH sensing. Yet, other titratable residues may account for residual proton sensitivity of H197A channels. A similar mutation at position 264 (H264A), alone or in combination with H197A, had no significant effect on PIP2;2 activity and pH sensitivity (data not shown). Loop D contains a stretch of charged amino acids that also may cooperate with the cytosolic pH-dependent function of His 197 (Fig. 4b, inset). Accordingly, the double mutants R194A/H197A and D195A/H197A displayed a complete insensitivity to cytosolic acidification at pH 6.0 (NaAc 6.0) (Fig. 4b). Single Ala mutants at positions 194 or 195 also showed a markedly reduced inhibition by pH (data not shown). Altogether, these results establish that charged residues in the cytosolic vestibule of the aqueous pore control its conformation and/or conductance and make these responsive to cytosolic protons. Although a His residue is lacking, mammalian aquaporins, including AQP1 and AQP2, also have a stretch of charged amino acids in loop D.

The structural basis described here for cytosolic pH-dependent gating is conserved in all PIPs. This provides a mechanism for coordinate inhibition of plant plasma membrane aquaporins and, as a consequence, for a general block of root water transport, as was observed at the cell and organ levels. Thus, cytosol acidosis was identified as the primary cause for inhibition of $L_{p,r}$ in anoxic conditions and represents, in addition to changes in cytosolic Ca^{2+} and Rop-dependent H_2O_2 production^{20,25}, a new signalling mechanism for anoxia. Furthermore, aquaporin mutants with altered sensitivity to cytosolic pH will be instrumental in testing how aquaporin activity in plant roots may be detrimental under anoxic stress, and this may provide new knowledge to improve the flooding tolerance of crops. □

Methods

Plant culture

Arabidopsis thaliana plants (ecotype Wassilewskija) were grown in hydroponic culture as described⁶.

Whole-root and cortex cell hydraulic conductivities

The pressure (P)-dependent flow (J_v) of sap exuded by excised roots was measured in a pressure chamber using the device described in ref. 6 and in Supplementary information 2, with a RBS containing 20 mM KCl, 1.25 mM KNO_3 , 1.5 mM $\text{Ca}(\text{NO}_3)_2$, 0.75 mM MgSO_4 , 0.05 mM KH_2PO_4 , 3 mM glucose, 10 mM Mes-KOH, pH 6.0. When indicated, J_v measurements were made in a RBS supplemented with 1 mM NaN_3 or 0.5 mM KCN or in a RBS in which 20 mM propionic acid/potassium propionate (KProp), pH 6.0, was substituted for 20 mM KCl. The dry weight (DW) of individual root systems was determined and hydrostatic hydraulic conductivity ($L_{p,r}$) was calculated from the slope of a J_v/DW to P plot derived from steady-state J_v measurements at five P values between 0.15 and 0.35 MPa.

Cell pressure probe measurements were performed as described⁶. Excised root tips (30–50 mm in length) were preincubated in a simplified RBS (KCl) containing 20 mM KCl, 10 mM Mes-KOH, pH 6.0, or in a corresponding acid load solution (KProp: 20 mM K-propionate, 10 mM Mes-KOH, pH 6.0), and the hydraulic conductivity of root cortex cells

($L_{p,\text{cell}}$) located between 2.5–3.5 mm from root apex was measured in the following 12–20 min.

In vivo NMR analyses

^{31}P -NMR spectra of excised roots (approximately 10 g fresh weight) were recorded at 20 °C in an AMX 400 wide bore spectrometer (Bruker) equipped with a 25-mm multinuclear probe tuned at 161.9 MHz, using an experimental arrangement described previously¹⁹. The acquisition of data used 70- μs pulses (50°) at 0.6-s intervals, and a sweep width of 9.8 kHz. A broadband decoupling at 2.5 W during acquisition and 0.5 W during the delay was applied using the Waltz sequence; the signal was digitized at 4,000 data points zero-filled to 8,000, and processed with a 5-Hz line broadening. Spectra were referenced to methylene diphosphonic acid (pH 8.9) as described¹⁹. Roots were continuously perfused at a rate of 120 ml min^{-1} with a RBS bubbled with O_2 or N_2 (anoxia) (gas flow of 1 l min^{-1}) in the medium reservoir. This device provides an optimal supply of O_2 in control conditions and allows one to evacuate the CO_2 produced by roots in anoxic conditions.

To assign the resonance of inorganic phosphate and soluble phosphate-containing compounds to specific peaks observed on *in vivo* ^{31}P -NMR spectra, the spectra of perchloric acid extracts prepared from the samples frozen immediately after *in vivo* analyses were compared with the spectra of standard solutions of known compounds¹⁹. The definitive assignments were made after running a series of spectra obtained by addition of the authentic compounds to the perchloric acid extracts at different pH values. Intracellular pH values were estimated from the chemical shift of the cytosolic and vacuolar Pi pools as previously described¹⁹. When the peak of cytosolic Pi was too weak to be unambiguously identified, cytosolic pH was evaluated from the chemical shift of the relatively abundant P-Cho (see Supplementary Information 1).

Expression in oocytes, water permeability and cytosolic pH

The complementary DNA of all aquaporins investigated was cloned in a pSP64T-derived expression vector²⁴. Aquaporin mutants were generated by oligonucleotide-directed mutagenesis using a recombinant PCR technique¹³ or the QuikChange Site-Directed Mutagenesis kit (Stratagene). All the constructs were verified by automated sequencing of the full cDNA. Complementary RNA production, expression in *Xenopus* oocytes and osmotic water permeability (P_f) measurements were performed as previously described¹³. Intracellular pH was measured with respect to a membrane potential electrode using a H^+ -selective liquid ion-exchanger microelectrode (hydrogen ionophore II Cocktail A, Fluka) connected to a high impedance amplifier as described²⁶. Electrodes had a linear response with a slope ≥ 52 mV pH unit⁻¹.

Received 18 February; accepted 5 June 2003; doi:10.1038/nature01853.

- Vartapetian, B. B. & Jackson, M. B. Plant adaptations to anaerobic stress. *Ann. Bot.* **79** (Suppl. A), 3–20 (1997).
- Clarkson, D. T. *et al.* Root hydraulic conductance: diurnal aquaporin expression and the effects of nutrient stress. *J. Exp. Bot.* **51**, 61–70 (2000).
- Birner, T. P. & Steudle, E. Effects of anaerobic conditions on water and solute relations, and on active transport in roots of maize (*Zea mays* L.). *Planta* **190**, 474–483 (1993).
- Kamaluddin, M. & Zwiazek, J. J. Metabolic inhibition of root water flow in red-osier dogwood (*Cornus stolonifera*) seedlings. *J. Exp. Bot.* **52**, 739–745 (2001).
- Zhang, W. H. & Tyerman, S. D. Effect of low O_2 concentration and azide on hydraulic conductivity and osmotic volume of the cortical cells of wheat roots. *Austr. J. Plant Physiol.* **18**, 603–613 (1991).
- Javot, H. *et al.* Role of a single aquaporin isoform in root water uptake. *Plant Cell* **15**, 509–522 (2003).
- Martre, P. *et al.* Plasma membrane aquaporins play a significant role during recovery from water deficit. *Plant Physiol.* **130**, 2101–2110 (2002).
- Siefritz, F., Tyree, M. T., Lovisolo, C., Schubert, A. & Kaldenhoff, R. PIP1 plasma membrane aquaporins in tobacco: from cellular effects to function in plants. *Plant Cell* **14**, 869–876 (2002).
- Zhang, W. H. & Tyerman, S. D. Inhibition of water channels by HgCl_2 in intact wheat root cells. *Plant Physiol.* **120**, 849–857 (1999).
- Parisi, M. & Bourguet, J. Effects of cellular acidification on ADH-induced intramembrane particle aggregates. *Am. J. Physiol.* **246**, C157–C159 (1984).
- Han, Z., Wax, M. B. & Patil, R. V. Regulation of Aquaporin-4 water channels by phorbol ester-dependent protein phosphorylation. *J. Biol. Chem.* **273**, 6001–6004 (1998).
- Johansson, I. *et al.* Water transport activity of the plasma membrane aquaporin PM28A is regulated by phosphorylation. *Plant Cell* **10**, 451–459 (1998).
- Maurel, C., Kado, R. T., Guern, J. & Chrispeels, M. J. Phosphorylation regulates the water channel activity of the seed-specific aquaporin α -TIP. *EMBO J.* **14**, 3028–3035 (1995).
- Yasui, M. *et al.* Rapid gating and anion permeability of an intracellular aquaporin. *Nature* **402**, 184–187 (1999).
- Zeuthen, T. & Klaefer, D. A. Transport of water and glycerol in Aquaporin 3 is gated by H^+ . *J. Biol. Chem.* **274**, 21631–21636 (1999).
- Németh-Cahalan, K. L. & Hall, J. E. pH and calcium regulate the water permeability of Aquaporin 0. *J. Biol. Chem.* **275**, 6777–6782 (2000).
- Gerbeau, P. *et al.* The water permeability of *Arabidopsis* plasma membrane is regulated by divalent cations and pH. *Plant J.* **30**, 71–81 (2002).
- Felle, H. H. pH: signal and messenger in plant cells. *Plant Biol.* **3**, 577–591 (2001).
- Gout, E., Boisson, A.-M., Aubert, S., Douce, R. & Bligny, R. Origin of the cytoplasmic pH changes during anaerobic stress in higher plant cells. Carbon-13 and phosphorus-31 nuclear magnetic resonance studies. *Plant Physiol.* **125**, 912–925 (2001).
- Subbiah, C. C., Bush, D. S. & Sachs, M. M. Elevation of cytosolic calcium precedes anoxic gene expression in maize suspension-cultured cells. *Plant Cell* **6**, 1747–1762 (1994).
- Steudle, E. & Peterson, C. A. How does water get through roots? *J. Exp. Bot.* **49**, 775–788 (1998).
- Javot, H. & Maurel, C. The role of aquaporins in root water uptake. *Ann. Bot.* **90**, 301–313 (2002).
- Tomos, A. D. & Leigh, R. A. The pressure probe: a versatile tool in plant cell physiology. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 447–472 (1999).
- Maurel, C., Reizer, J., Schroeder, J. I. & Chrispeels, M. J. The vacuolar membrane protein γ -TIP creates water specific channels in *Xenopus* oocytes. *EMBO J.* **12**, 2241–2247 (1993).

25. Baxter-Burrell, A., Yang, Z., Springer, P. S. & Bailey-Serres, J. RopGAP4-dependent Rop GTPase rheostat control of *Arabidopsis* oxygen deprivation tolerance. *Science* **296**, 2026–2028 (2002).
26. Lacombe, B., Pilot, G., Gaymard, F., Sentenac, H. & Thibaud, J.-B. pH control of the plant outwardly-rectifying potassium channel SKOR. *FEBS Lett.* **466**, 351–354 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank H. Höfte, H. Sentenac and C. Vander Willigen for critical reading of the manuscript; N. Declercq for discussions; Y. Boursiac for help with cell pressure probe measurements; and J. B. Thibaud and C. Plassard for assistance in micro-electrode pH measurements. This work was supported in part by the Centre National de la Recherche Scientifique (Action Thématique Incitative sur Programme et Equipe 'Function and regulation of plant aquaporins'). M.S. is on leave from the Laboratorio de Biomembranas, Departamento de Fisiología, Facultad de Medicina, Universidad de Buenos Aires, Argentina.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.M. (maurel@ensam.inra.fr).

ER-phagosome fusion defines an MHC class I cross-presentation compartment in dendritic cells

Pierre Guermonprez^{1*}, Loredana Saveanu^{2*}, Monique Kleijmeer³, Jean Davoust⁴, Peter van Endert² & Sebastian Amigorena¹

¹INSERM U520, Institut Curie, 26 rue d'Ulm, 75005 Paris, France

²INSERM U580, Institut Necker, 161 rue de Sévres, 75015 Paris, France

³Department of Cell Biology, UMC Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

⁴CNRS UMR8115, Genethon, 1 bis rue de l'Internationale, 91002 Cedex 02, Evry, France

* These authors contributed equally to this work

Induction of cytotoxic T-cell immunity requires the phagocytosis of pathogens, virus-infected or dead tumour cells by dendritic cells¹. Peptides derived from phagocytosed antigens are then presented to CD8⁺ T lymphocytes on major histocompatibility complex (MHC) class I molecules, a process called "cross-presentation"^{2,3}. After phagocytosis, antigens are exported into the cytosol and degraded by the proteasome^{4–6}. The resulting peptides are thought to be translocated into the lumen of the endoplasmic reticulum (ER) by specific transporters associated with antigen presentation (TAP), and loaded onto MHC class I molecules by a complex "loading machinery" (which includes tapasin, calreticulin and Erp57)⁷. Here we show that soon after or during formation, phagosomes fuse with the ER. After antigen export to the cytosol and degradation by the proteasome, peptides are translocated by TAP into the lumen of the same phagosomes, before loading on phagosomal MHC class I molecules. Therefore, cross-presentation in dendritic cells occurs in a specialized, self-sufficient, ER-phagosome mix compartment.

The reasons why phagocytosis itself favours cross-presentation in dendritic cells and macrophages are unclear. To gain further insight into this process, we purified phagosomes from immature dendritic cells. After purification, the enrichment of various phagosomal markers was analysed by western blotting. Early phagocytic markers, such as transferring receptor and phosphoinositide-3-kinase (Pi3Kp85), were detected early after engulfment and rapidly decreased thereafter. Markers of the late endocytic pathway, such as rab7 (Fig. 1a) and Lamp2 (Fig. 1b, left panels) increased over time, indicating phagosome maturation towards phagolysosomes.

In macrophages, phagosomes were recently shown to fuse with the ER during or soon after engulfment⁸. Similarly, we found that several ER resident proteins such as sec61/sec62 and calnexin were detected immediately after phagosome formation in dendritic cells (Fig. 1a and Fig. 1b, right panels). Recruitment of ER residents into phagosomes, however, was restricted to a subset of the ER proteins detected by polyclonal antibodies raised against whole ER membranes (Supplementary Information SD1). The presence of these ER markers in phagosomes decreased as phagosomes matured into phagolysosomes.

To quantify the proportion of phagosomes bearing ER markers, we used flow cytometric (FACS) analysis of phagosomes⁹. After the pulse with latex beads, only 10% of the phagosomes expressed Lamp2, and the proportion of Lamp2-bearing phagosomes increased over time (Fig. 1b, left panel), to reach over 50% after 240 min chase. The distribution of calnexin, in contrast, was homogeneously high in early phagosomes (right panels in Fig. 1b), suggesting that most of the phagosomes had already fused with ER membranes. Consistent with our biochemical analysis and with previous results in macrophages⁸, calnexin labelling of phagosomes decreased over time.

ER recruitment to phagosomes was also shown in whole cells, using confocal-laser scanning-microscopy (CLSM). In immature dendritic cells pulsed with fluorescent beads, staining of the phagosomal membrane was observed using both total ER antibodies (not shown) and calnexin antibodies (Fig. 1e), in a wide proportion of early phagosomes. Electronic microscopy analysis of ER distribution by calreticulin staining demonstrates the juxtaposition of ER membranes to the forming phagosomes (phagocytic cups were stabilized by PI3K inhibitors⁸), and strongly suggests direct fusion between the two organelles (Fig. 1c). Direct evidence for ER-phagosome fusion was obtained using a cytochemical EM technique, revealing the activity of the ER enzyme glucose-6-phosphatase (G-6-Pase) on Epon sections¹⁰. G-6-Pase activity appears as dense precipitates in numerous ER structures throughout the cells, including the nuclear envelope (Fig. 1d). Strikingly, the majority of phagosomes displayed strong G-6-Pase labelling sometimes over the whole of the limiting membrane (Fig. 1d, right panel), other times on part of it (Fig. 1d, left panel). We concluded that in dendritic cells, like in macrophages⁸, phagosomes fuse with the ER, soon after or during particle engulfment. ER proteins then progressively disappear as phagosomes mature into phagolysosomes.

Fusion with the ER and the recruitment of ER proteins to phagosomes led us to reconsider the conventional view of cross-presentation. We hypothesized that after antigen export to the cytosol and degradation by the proteasome, peptides could be translocated back into the lumen of the same phagosome by TAP, and loaded onto MHC class I molecules in the phagosomal lumen.

To test this model, we first examined whether ER fusion with phagosomes actually results in recruitment of TAP and of the MHC class I loading machinery. Abundant TAP was detected in early phagosomes and was reduced over time (Fig. 2a, right panel). Flow cytometric analysis of phagosome populations⁹ of various ages revealed homogeneous staining of early phagosomes, which also decreased over time (Fig. 2a, left panel). EM analysis of purified phagosomes showed that TAP2 is inserted into the phagosomal membrane surrounding the latex bead (Fig. 2b). The recruitment of TAP to phagosomes was also seen in intact dendritic cells using CLSM (Fig. 2c). In spite of high staining of abundant ER membranes throughout the cells, enrichment in membranes surrounding the beads was evident in most phagosomes.

Together with TAP, tapasin, calreticulin, Erp57 and the heavy chain of MHC class I were all detected in purified early phagosomes by western blotting (Fig. 2d). A clear decrease over time was observed for tapasin and Erp57, whereas MHC class I heavy chain and calreticulin remained stable. Therefore, in the first few

hours after engulfment, phagosomes recruit the MHC class I peptide translocation and loading machinery.

We next examined whether cross-presentation of phagocytosed antigens occurs in the same time window as ER recruitment to phagosomes; that is, in the first 2–3 h after phagocytosis. The egress

of antigens from phagosomal lumen to the cytosol is detected after 1.5 h (Supplementary Information SD2). Consistently, cross-presentation of the OVA_{257–264} peptide (SIINFEKL) by the H-2K^b class I MHC molecule reached a plateau within 3 h of chase (Fig. 3a). At these early time points, cross-presentation of OVA-coated

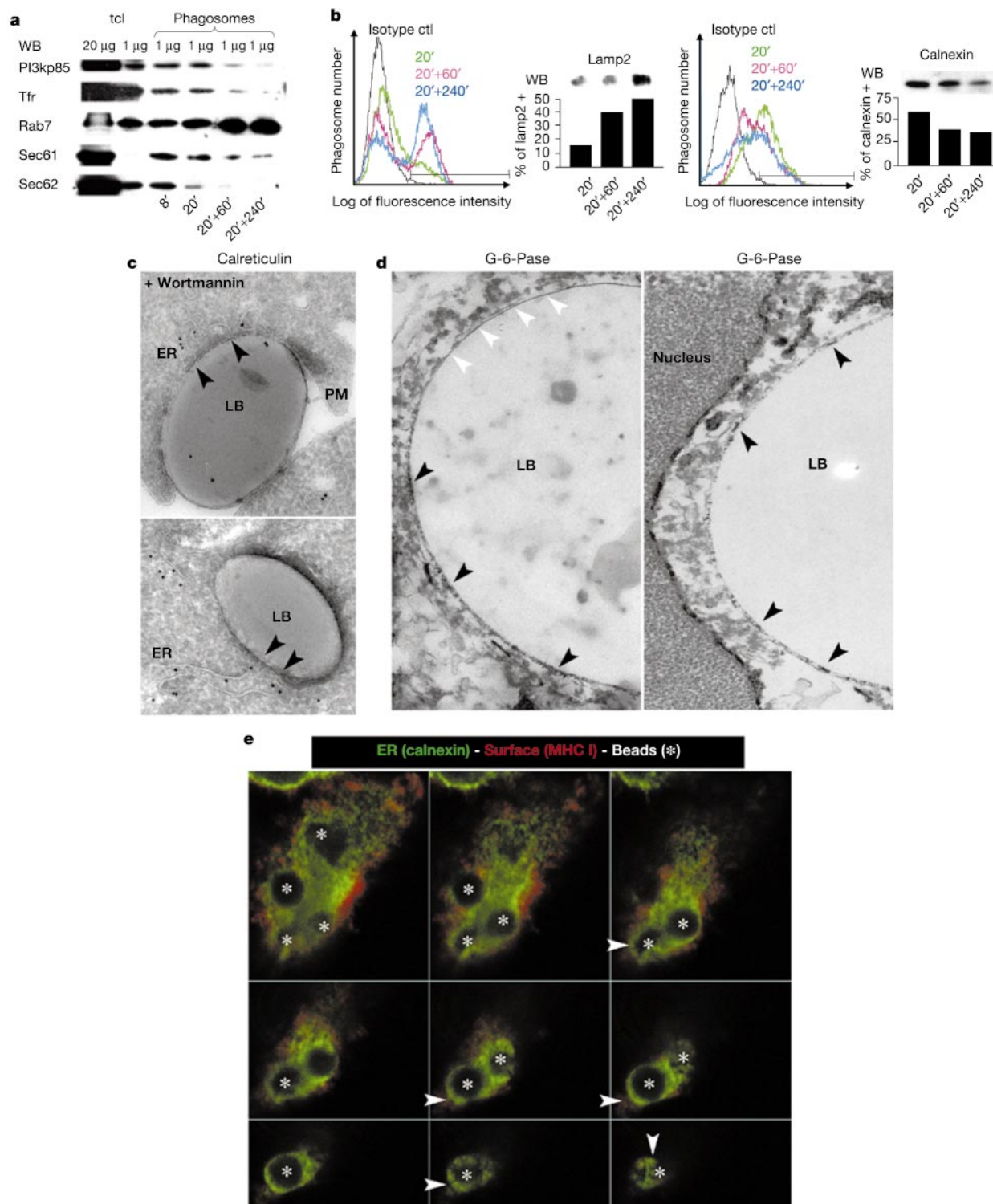


Figure 1 Rapid recruitment of ER at the early stages of phagosome biogenesis in immature dendritic cells. **a**, Western blot analysis of phagosomal maturation. tcl, total cell lysate. **b**, FACS and western blot (WB) analysis of calnexin and Lamp2 recruitment on phagosomes. **c**, Cryoelectron microscopy analysis of ER-phagosome fusion. Calreticulin-positive (gold beads) ER membranes are observed in direct apposition to the latex beads

(LB) phagosomes, suggesting fusion of the two compartments (arrows). **d**, EM G-6-Pase staining of ER and phagosomal membranes. Phagosomal membranes containing (black arrows) or not (white arrows) G-6-Pase precipitates. **e**, CLSM analysis of ER recruitment at early stages of phagosome biogenesis. Asterisks indicate beads, arrows indicate the accumulation of calnexin staining around beads. Ctrl, control.

beads by dendritic cells required TAP and was blocked by lactacystin, a potent proteasome inhibitor (Supplementary Information SD3).

These results raise the possibility that mixed ER-phagosome compartments represent new specialized organelles for loading of MHC class I molecules during cross-presentation.

Might peptide loading during cross-presentation occur in phagosomes? If so, at these early time points the complexes should accumulate phagosomes. We first analysed the subcellular distribution of MHC class I-peptide complexes using a monoclonal antibody specific for K^b molecules loaded with the OVA₂₅₇₋₂₆₄ peptide (K^b -OVA)¹¹. After 2 h of chase, cells that have phagocytosed BSA- or OVA beads were analysed by CLSM. As shown in Fig. 3b, abundant labelling was observed on the limiting phagosomal membrane as well as on punctuated structures in the vicinity of phagosomes containing OVA-beads, but not of phagosomes containing BSA-beads. Nine per cent of the OVA-phagosomes (1% of BSA-phagosomes) displayed continuous labelling of their limiting membrane on over two-thirds of the phagosomes (Fig. 3c). Over 30% of the OVA-phagosomes (2% of BSA-phagosomes) showed labelling on lower proportions of their limiting membrane. We concluded that K^b -OVA complexes accumulate in early OVA-phagosomes, suggesting that peptide loading may occur therein.

The accumulation of K^b -OVA complexes in phagosomes during cross-presentation was also examined using a functional assay for T-cell activation. OVA- or BSA-phagosomes were purified as before, and phagosomal membranes were then disrupted and separated from the beads. As shown in Fig. 3d, membranes from OVA-phagosomes, but not from BSA-phagosomes, induced IFN- γ secretion by OVA-specific OT-1 T cells. T-cell activation was blocked when the OVA-bead phagosomes were purified from lactacystin-treated cells (Fig. 3d). Therefore, during cross-presentation, MHC class I molecules loaded with peptides generated in the cytosol are present in phagosomes.

The presence of K^b -OVA complexes in early phagosomes could, however, result from either peptide loading in phagosomes, or from peptide loading elsewhere in the cell and redistribution to phagosomes. To distinguish between these two possibilities, we first developed a FACS-based technique to analyse K^b -OVA complexes quantitatively in individual phagosomes. D1 cells were pulsed with OVA-beads for 30 min and then chased for 2 h, before disruption of the cells. After fixation and permeabilization, the phagosomes were stained indirectly with the K^b -OVA-specific antibody and the fluorescence was analysed by FACS after gating on single-bead-containing phagosomes⁹. As shown in Fig. 3e, specific staining of the phagosomes was observed, as compared to the matched isotype control and to naked OVA-beads.

We reasoned that if phagosomes behaved as autonomous cross-presentation organelles (if they were both the source of antigen and the site of peptide loading), then bystander phagosomes, which do not contain OVA, but are present in OVA-cross-presenting cells, should not accumulate K^b -OVA complexes. Alternatively, if K^b -OVA complexes formed in the ER, and were subsequently transported to phagosomes, then both OVA- and BSA-phagosomes should contain similar amounts of K^b -OVA complexes. We therefore performed the following experiment: dendritic cells were incubated with a mixture of fluorescein isothiocyanate (FITC)-BSA-beads and OVA-beads, under conditions where over 80% of the cells have phagocytosed both types of beads. We then analysed the presence of K^b -OVA complexes in the two populations of phagosomes using the same K^b -OVA-specific antibody and FACS after gating on OVA (FITC-negative) or BSA (FITC-highly positive).

As shown in Fig. 3f, FITC-negative, OVA-containing phagosomes showed reproducibly stronger labelling with the K^b -OVA-specific antibody, as compared to BSA-containing phagosomes from the

same cells. The very low staining of BSA-phagosomes in the mixture was not antigen-specific, because it was similar to the staining obtained with BSA-phagosomes alone (Fig. 3f, g, cells fed with BSA-beads in the absence of OVA-beads). Because the fluorescence intensity distributions are unimodal, we compared the K^b -OVA and the isotype control stainings using the Kolmogorov-Smirnov statistical analysis on 25,000 phagosomes from each sample. As shown in Fig. 3g, the differences are statistically significant. Phagosomes therefore represent autonomous antigen cross-presentation compartments.

The results presented thus far support a model of phagosome-mediated cross-presentation, in which phagosomes can be a site of

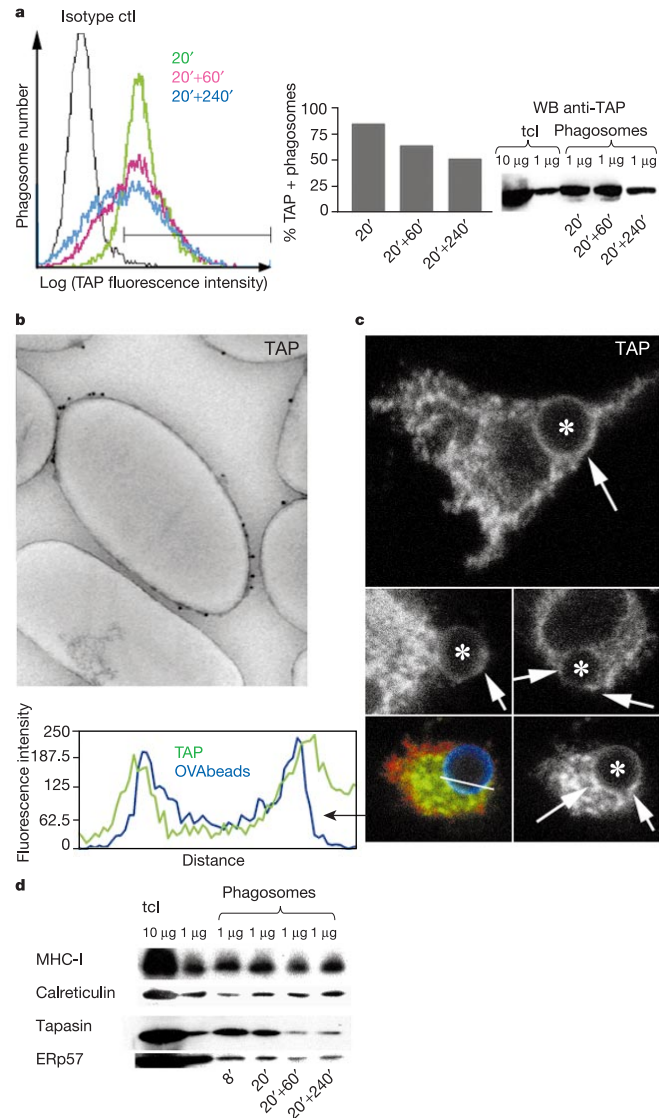


Figure 2 TAP is recruited in early phagosomes. **a**, FACS and WB analysis of TAP recruitment during phagosomal maturation. **b**, EM analysis of TAP recruitment on purified phagosomes. Purified 30 min, 0.8 µm latex bead-phagosomes were stained with a monoclonal antibody against the cytosolic portion of TAP2 (or an isotype control, not shown). **c**, CLSM analysis of TAP recruitment at early stages of phagosome biogenesis. Single confocal section of D1 cells pulsed for 30 min with 3 µm OVA-beads (asterisks) and stained for TAP2. Arrows indicate TAP enrichment in ruffles surrounding the forming phagocytic cups. For one cup, staining of cell surface MHC class I (in red) and the bead (blue) are shown (lower left panel). Fluorescence intensity along the white line (bottom left) was quantified for OVA at the surface of the latex bead (blue) and TAP2 (green). **d**, Recruitment of MHC class I and its loading machinery during phagosome biogenesis were analysed by western blot.

TAP-dependent peptide loading on MHC class I molecules. To directly test this model, we next attempted to reconstitute TAP-dependent peptide translocation and loading *in vitro*, using purified phagosomes. We reasoned that after fusion of ER membranes to phagosomes, the latter should acquire the N-glycosylation machinery, which can be used as a translocation readout in TAP-dependent transport assays. Purified phagosomes were incubated with a TAP substrate iodinated peptide (R-10-F*), bearing a sequence for N-glycosylation. After lysis, N-glycosylated R-10-F* was precipitated using ConA. We observed time- and ATP-dependent accumulation of N-glycosylated R-10-F*, indicating transport of R-10-F* into the lumen of phagosomes (Fig. 4a). Glycosylation of R-10-F* was competed by unlabelled R-10-F peptide (Fig. 4b). A cold TAP-substrate (peptide 435), devoid of N-glycosylation sites, competed with the glycosylation of R-10-F* (Fig. 4b), indicating that TAP was involved in peptide translocation across the phagosomal membrane.

The role of TAP in this process was directly shown by the absence of R-10-F* glycosylation in phagosomes isolated from bone marrow dendritic cells (BM-DCs) originating from TAP^{-/-} mice, but not from their TAP^{+/+} counterparts (Fig. 4c). TAP activity was not due

to contaminating ER membranes that associated to phagosomes after cell homogenization, in the course of the purification procedure, because addition of unpulsed TAP^{+/+} dendritic cells to latex-beads-pulsed TAP^{-/-} dendritic cells before purification of phagosomes did not correct the failure of the resulting TAP^{-/-} purified phagosomes to import and glycosylate R-10-F* peptide (Fig. 4d). In agreement with western blot and FACS analysis (Fig. 2), TAP-mediated import and glycosylation of R-10-F* decreased with phagosomal maturation (Fig. 4e). We concluded that dendritic cell phagosomes are competent for TAP-dependent peptide translocation during the first few hours after phagocytosis.

Can TAP-mediated import of cytosolic peptides to phagosomal lumen be readily coupled to phagosomal class I MHC loading? To address this question, we purified phagosomes from dendritic cells derived from mice expressing HLA-A2¹². The purified phagosomes were incubated with iodinated, HLA-A2-restricted, S-9-L*. Peptide loading was detected after immunoprecipitation with HLA-A2-specific antibodies, but not with isotype-matched control antibodies (Fig. 4f). Effective loading required the presence of ATP (Fig. 4f), suggesting that TAP-mediated transport was required before loading. Loading was competed by an excess of unlabelled

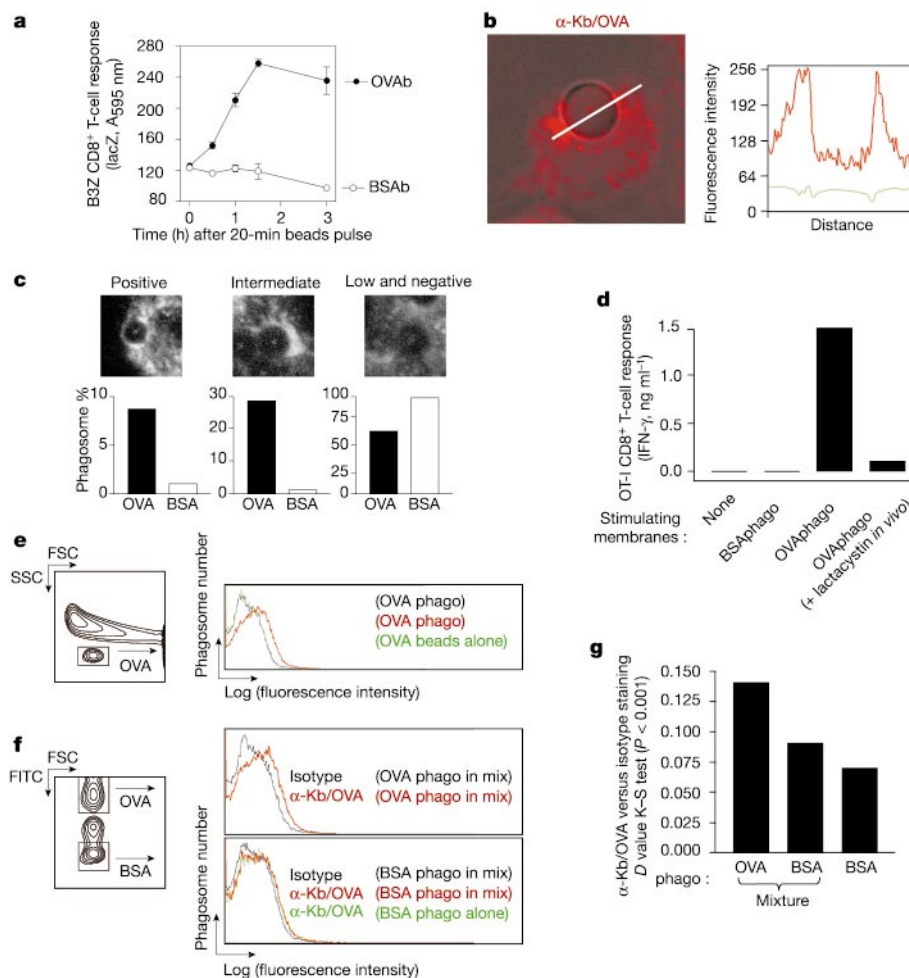


Figure 3 Early peptide loading in phagosomes during cross-presentation. **a**, Kinetics of OVA cross-presentation after phagocytosis. **b**, Intracellular localization of H-2K^b/OVA complexes to phagosomes in cross-presenting dendritic cells using 25-D1.16 antibody. An image of an OVA-containing phagosome is shown (overlay of transmission image (grey) and 25-D1.16 (red)). A linescan of transmission (green) and fluorescence intensity of 25-D1.16 (red) are shown in the right panel. **c**, Quantification of H-2K^b/OVA staining in OVA- and BSA-phagosomes. 100 of each BSA- or OVA-phagosomes were randomly scored for 25-D1.16 staining. **d**, Functional H-2K^b/OVA complexes formed after

proteasomal degradation are present in purified early phagosomes. **e**, FACS detection of H-2K^b/OVA in individual early phagosomes using 25-D1.16 antibody. **f**, Phagosomes represent autonomous compartments for cross-presentation. OVA-phagosomes accumulate H-2K^b/OVA, as compared to BSA-phagosomes from the same cells (or to BSA-phagosomes from cells that did not capture OVA-beads). **g**, Kolmogorov-Smirnov (K-S) statistical analysis of the H-2K^b/OVA stainings from **f**. *D* value is the difference between K^b/OVA and isotype stainings.

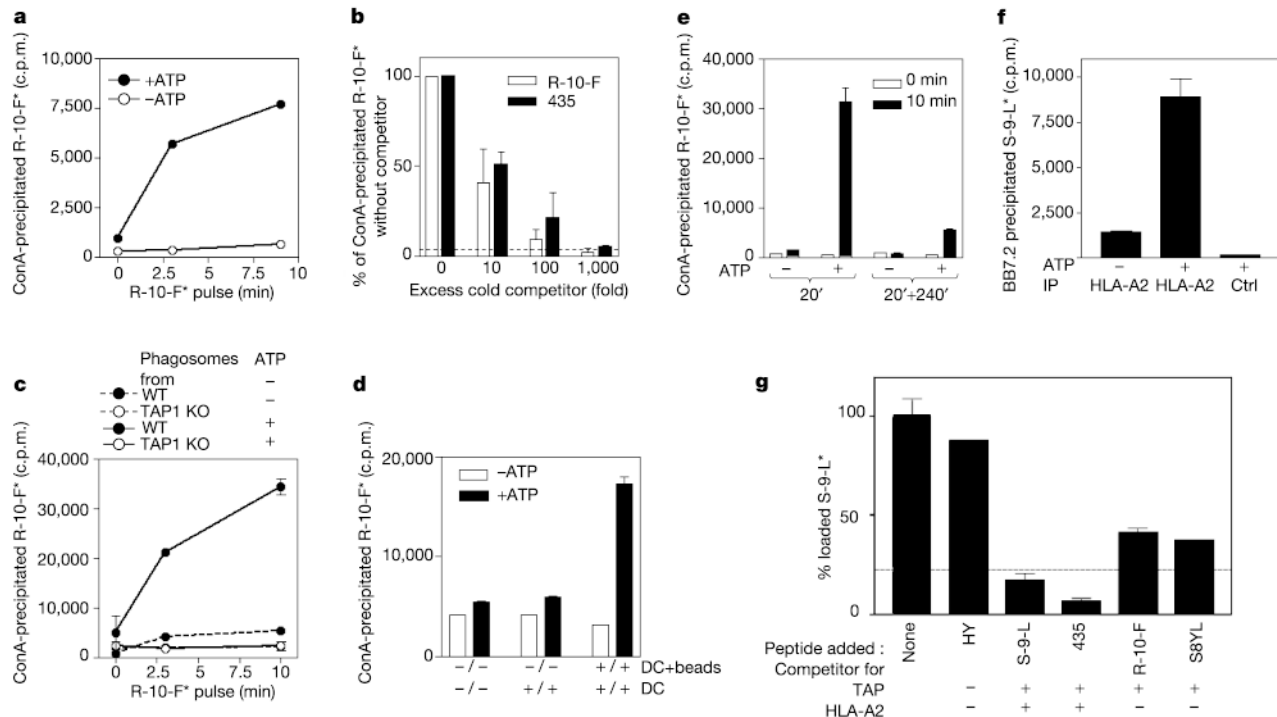


Figure 4 *In vitro* reconstitution of TAP-mediated peptide transport and loading on MHC class I in purified phagosomes. **a**, Time and ATP-dependent *n*-glycosylation of the R-10-F* peptide. 30 min phagosomes from D1 cells were incubated for indicated times with the iodinated R-10-F*, in the presence or not of ATP. Glycosylated R-10-F* measured using ConA-sepharose. c.p.m., cycles per minute. **b**, R-10-F* import and luminal glycosylation are competed by 435, a TAP peptide substrate devoid of *n*-glycosylation site. **c**, Phagosomes purified from TAP1^{-/-} BMDCs do not exhibit ATP-dependent import of R-10-F* peptide for luminal *n*-glycosylation. WT, wild type; KO, knock out. **d**, Accumulation of glycosylated peptide R-10-F* in phagosomes is not due to *in vitro* ER contamination during phagosome purification. BMDCs (TAP1^{+/+} or ^{-/-}) pulsed with

beads were mixed with unpulsed cells (TAP1^{+/+} or ^{-/-}) at 4 °C, before phagosome purification. **e**, Accumulation of glycosylated R-10-F* is reduced upon phagosome maturation. **f**, *In vitro* reconstitution of ATP-dependent loading of MHC class I in purified early phagosomes. 30 min phagosomes purified from HLA-A2 transgenic mice BMDCs were incubated with iodinated S-9-L*, an HLA-A2 binding peptide, in the presence or absence of ATP. Loading of S-9-L* on HLA-A2 was measured after immunoprecipitation (IP). **g**, ATP-dependent loading of MHC class I in purified phagosomes is competed by peptides with high TAP affinity that do not bind to MHC class I. The dotted line indicates the average threshold for ATP-independent background binding.

S-9-L* and by 435, an unrelated HLA-A2-binding peptide (Fig. 4g). Moreover, loading of S-9-L* on HLA-A2 was also blocked using cold TAP-substrate peptides that did not bind to HLA-A2: both R-10-F and S-8-YL nearly completely inhibited ATP-dependent S-9-L* peptide loading on HLA-A2 (Fig. 4g). The specificity of the inhibitions was confirmed using an irrelevant peptide, H-Y, which has a low affinity for murine TAP, and does not bind HLA-A2. We concluded that TAP-imported peptides can be loaded on MHC class I molecules in the lumen of phagosomes.

We show here that ER-phagosome fusion defines a compartment for proteasome- and TAP-dependent cross-presentation in phagosomes. Mix ER-phagosome compartments bring together all the MHC class I-processing²² and loading machinery, as well as exogenous antigens, in a single subcellular organelle. By doing so, antigen-presenting cells cope with the paucity of exogenous antigens that could gain access to proteasomal degradation (as compared to the abundant and sustained supply of newly synthesized proteins), the low efficiency of antigenic peptide generation by proteasomes¹³ and the short half-life of these antigenic peptides in the cytosol¹⁴. By limiting the competition with endogenous antigens, dendritic cells may focus all the cross-presentation machinery on the antigens relevant for the initiation of most immune responses, that is, those acquired by phagocytosis. □

Methods

Mice and cells

TAP1^{+/+} (C57BL/6 WT) or TAP1^{-/-} (TAP1^{-/-}, li^{-/-}, backcrossed on a B/6 background, provided by O. Lantz), HLA-A2 transgenic mice (H-2K^b-/-, H-2D^b-/-, murine82m^{-/-},

human82m^{+/+}, HLA-A2^{+/+} (ref. 12), provided by F. Lemonnier), OT-1 (H-2K^b restricted, anti-OVA TCR transgenic, RAG1^{-/-} mice¹⁵, provided by B. Combadière) were used in this study. The immature splenic dendritic cell line D1 was provided by P. Ricciardi-Castagnoli¹⁶. BMDCs were obtained from mice after 8–10 days culture of BM cells in GM-CSF-containing medium as described in ref. 17.

Antibodies

Anti-PI3Kp85 (Upstate), anti-TfR H68.4 (ATCC), anti-Rab7 (J.-P. Gorvel), purified anti-Sec61 and anti-Sec62 (E. Hartmann), anti-Lamp2 ABL-93 (Pharmingen), anti-calnexin SPA-860 and SPA-865 (Stressgen), anti-whole ER membranes rabbit serum (D. Louvard), anti-TAP2 435, anti-TAP2 (J. Monaco), anti-MHC class I P8 (H. Ploegh), anti-calreticulin PA3-900 (Affinity), anti-tapasin 5D3 (H. Hansen), anti-Erp57 SPA-585 (StressGen) and anti-K^b/OVA 25-D1.16 (R. Germain) antibodies were used. Anti-species conjugated to Alexa-488 (Molecular Probes), Texas-Red (Jackson Laboratories) or peroxidase (Jackson Laboratories) were used as second-step reagents.

Phagosome purification

For phagosome purification, cells were pulsed/chased with 0.8 μm latex beads (LB, Sigma) and phagosomes were purified according to refs 8, 18. Phagosomes containing LB were collected at the 25–8.5% interface of a discontinuous sucrose gradient and washed in Transport Buffer (KCl 130 mM, 10 mM NaCl, 1 mM CaCl₂, 2 mM EGTA, 1 mM DTT, 5 mM Hepes, pH 7.3). Then they were lysed in NP40 buffer (NP40 0.5%, MgCl₂ 5 mM, Tris 50 mM pH 7.4). Protein content was quantified using a modified Lowry assay (Biorad). Purified phagosomes were used in transport assay or were submitted to SDS-PAGE in reducing conditions.

FACS analysis of phagosomes

Phagosomes from homogenized cells were analysed by FACS according to the method described⁹. Preparations were analysed after gating on a particular FSC/SSC region (corresponding to single beads in a beads solution), which allows the distinction of single-bead-containing phagosomes from other organelles. See Supplementary Information SD4.

CLSM

For CLSM, cells pulsed with 3 µm LB (Polyscience, coated with OVA-FITC conjugates) were allowed to adhere on polylysine-coated coverslips. Before fixation (4% paraformaldehyde-PBS, 10 min), cells were permeabilized before staining (PBS-saponin 0.2%, 10 min; except for Fig. 1a, class I MHC staining was performed before permeabilization in order to selectively stain cell surface with the anti H-2K^b Y3 antibody) and stained and washed in PBS BSA 0.1% saponin 0.1%. After mounting, coverslips were analysed on a Leica SP2 confocal microscope.

In vitro TAP and loading assays

A TAP transport assay was performed¹⁹. The same procedure was used for HLA-A2 loading assay, except that HLA-A2 bound peptide was immunoprecipitated on anti-HLA-A2- or control Sepharose. See Supplementary Information SD4.

Cross-presentation assays

3 µm LB were coated with OVA (or BSA) protein by passive adsorption (incubation in 10 mg ml⁻¹ PBS solutions for 18 h at 4 °C) and extensively washed in PBS. DCs were pulsed and chased for indicated times with OVA-beads or BSA-beads. After fixation (0.08% glutaraldehyde 3 min, stopped in glycine 0.2 M) cells were washed and put in culture (10⁵ per well in 96 flat well plates) for 18 h with the B3Z CD8⁺ T-cell hybridoma specific for the H-2K^b/OVA_{257–264} complex (10⁵ per well)²⁰. B3Z activation was monitored by the induction of lacZ reporter under NF-AT elements using the CPRG substrate (Roche)²⁰. For lactacystin inhibition assays, DCs were preincubated for 20 min before beads pulse with 18 µM lactacystin and lactacystin was kept during the entire pulse period.

T-cell activation using purified phagosomal membranes

D1 cells were pretreated or not with 18 µM lactacystin and pulsed with OVA-beads or BSA-beads for 30 min. After a 2-h chase (in the continuous presence of lactacystin or not), the cells were homogenized and the phagosomes were purified as before. The phagosomal membranes were then disrupted by freeze–thaw, separated from the beads by low-speed centrifugation and pelleted by ultracentrifugation (100,000g, 1 h). Anti-H-2K^b/OVA transgenic OT-1 effector T cells (obtained by *in vitro* stimulation with peptide-pulsed DCs and culture for 3 to 15 days in the presence of IL-2) were then added and IFN-γ secretion was measured after an 18-h incubation in supernatants with an ELISA test (Pharmingen).

Cryoelectron microscopy and G-6-Pase staining

D1 cells were incubated with 0.8 µm LB for 1 h (Fig. 1c) or 3 µm LB for 30 min (Fig. 1d), before fixation in 2% paraformaldehyde, 0.2% glutaraldehyde. Then cells were processed for ultrathin cryosectioning and immunogold labelling as described²¹. For G-6-Pase staining, D1 cells were incubated with 3 µm LB for 1 h and the Epon sections were processed as previously described¹⁰. When indicated, the cells pretreated with 100 nM wortmannin and then pulsed for 1 h with the LB in the continuous presence of the drug.

Received 17 March; accepted 10 July 2003; doi:10.1038/nature01911.

- Guermonprez, P., Valladeau, J., Zitvogel, L. & Amigorena, S. Antigen Presentation and T Cell Stimulation by Dendritic Cells. *Annu. Rev. Immunol.* **20**, 620–667 (2002).
- Heath, W. R. & Carbone, F. R. Cross-presentation, dendritic cells, tolerance and immunity. *Annu. Rev. Immunol.* **19**, 47–64 (2001).
- Yewdell, J. W., Norbury, C. C. & Bennink, J. R. Mechanisms of exogenous antigen presentation by MHC class I molecules *in vitro* and *in vivo*: implications for generating CD8⁺ T cell responses to infectious agents, tumors, transplants, and vaccines. *Adv. Immunol.* **73**, 1–77 (1999).
- Kovacsics-Bankowski, M. & Rock, K. L. A phagosome-to-cytosol pathway for exogenous antigens presented on MHC class I molecules. *Science* **267**, 243–246 (1995).
- Norbury, C. C., Chambers, B. J., Prescott, A. R., Ljunggren, H. G. & Watts, C. Constitutive macropinocytosis allows TAP-dependent major histocompatibility complex class I presentation of exogenous soluble antigen by bone marrow-derived dendritic cells. *Eur. J. Immunol.* **27**, 280–288 (1997).
- Rodriguez, A., Regnault, A., Kleijmeer, M., Ricciardi-Castagnoli, P. & Amigorena, S. Selective transport of internalized antigens to the cytosol for MHC class I presentation in dendritic cells. *Nature Cell Biol.* **1**, 362–368 (1999).
- Cresswell, P., Bangia, N., Dick, T. & Diedrich, G. The nature of the MHC class I peptide loading complex. *Immunol. Rev.* **172**, 21–28 (1999).
- Gagnon, E. *et al.* Endoplasmic reticulum-mediated phagocytosis is a mechanism of entry into macrophages. *Cell* **110**, 119–131 (2002).
- Ramachandra, L., Sramkoski, R. M., Canaday, D. H., Boom, W. H. & Harding, C. V. Flow analysis of MHC molecules and other membrane proteins in isolated phagosomes. *J. Immunol. Methods* **213**, 53–71 (1998).
- Griffiths, G., Quinn, P. & Warren, G. Dissection of the Golgi complex. I. Monensin inhibits the transport of viral membrane proteins from medial to trans Golgi cisternae in baby hamster kidney cells infected with Semliki Forest virus. *J. Cell Biol.* **96**, 835–850 (1983).
- Porgador, A., Yewdell, J. W., Deng, Y., Bennink, J. R. & Germain, R. N. Localization, quantitation, and *in situ* detection of specific peptide-MHC class I complexes using a monoclonal antibody. *Immunity* **6**, 715–726 (1997).
- Lemonnier, F. A. The utility of H-2 class I knockout mice. *Virus Res.* **82**, 87–90 (2002).
- Rock, K. L., York, I. A., Saric, T. & Goldberg, A. L. Protein degradation and the generation of MHC class I-presented peptides. *Adv. Immunol.* **80**, 1–70 (2002).
- Reits, E. *et al.* Peptide diffusion, protection, and degradation in nuclear and cytoplasmic compartments before antigen presentation by MHC class I. *Immunity* **18**, 97–108 (2003).
- Hogquist, K. A. *et al.* T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17–27 (1994).

- Winzler, C. *et al.* Maturation stages of mouse dendritic cells in growth factor-dependent long-term cultures. *J. Exp. Med.* **185**, 317–328 (1997).
- Thery, C. *et al.* Molecular characterization of dendritic cell-derived exosomes: selective accumulation of the heat shock protein hsc73. *J. Cell Biol.* **147**, 599–610 (1999).
- Desjardins, M., Huber, L. A., Parton, R. G. & Griffiths, G. Biogenesis of phagolysosomes proceeds through a sequential series of interactions with the endocytic apparatus. *J. Cell Biol.* **124**, 677–688 (1994).
- Daniel, S., Caillaud-Zucman, S., Hammer, J., Bach, J. F. & van Ender, P. M. Absence of functional relevance of human transporter associated with antigen processing polymorphism for peptide selection. *J. Immunol.* **159**, 2350–2357 (1997).
- Karttunen, J., Sanderson, S. & Shastri, N. Detection of rare antigen-presenting cells by the lacZ T-cell activation assay suggests an expression cloning strategy for T-cell antigens. *Proc. Natl Acad. Sci. USA* **89**, 6020–6024 (1992).
- Raposo, G., Kleijmeer, M. J., Posthuma, G., Slot, J. W. & Geuze, H. J. in *Handbook of Experimental Immunology* 5th edn, Ch. 208, 1–11 (Elsevier, Cambridge, MA, 1997).
- Houde, M. *et al.* Phagosomes are competent organelles for antigen cross-presentation. *Nature* **425**, 402–406 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We want to thank S. Hugues for help with the OT-1 effector cells, J. Griffith for technical assistance and A. Lennon-Dusmenil for critical reading of the manuscript. P.G. is a Fellow of the ARC and CNRS. This work was supported by the Institut Curie, the INSERM and the Ligue de Contre le Cancer.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.A. (sebastian.amigorena@curie.fr).

Phagosomes are competent organelles for antigen cross-presentation

Mathieu Houde^{1*}, Sylvie Bertholet^{2*}, Etienne Gagnon¹, Sylvain Brunet¹, Guillaume Goyette¹, Annie Laplante¹, Michael F. Princiotta³, Pierre Thibault⁴, David Sacks² & Michel Desjardins^{1,4}

¹Département de pathologie et biologie cellulaire, Université de Montréal, C.P.6128, Succ centre-ville, Montréal, Québec, H3C 3J7, Canada

²Laboratory of Parasitic Diseases, and ³Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

⁴Caprion Pharmaceuticals Inc., Montréal, Québec, H4S 2C8, Canada

* These authors contributed equally to this work

The ability to process microbial antigens and present them at the surface of cells is an important aspect of our innate ability to clear infections. It is generally accepted that antigens in the cytoplasm are loaded in the endoplasmic reticulum and presented at the cell surface on major histocompatibility complex (MHC) class I molecules, whereas peptides present in endo/phagocytic compartments are presented on MHC class II molecules^{1,2}. Despite the apparent segregation of the class I and class II pathways, antigens from intracellular pathogens including mycobacteria, *Escherichia coli*, *Salmonella typhimurium*, *Brucella abortus* and *Leishmania*, have been shown to elicit an MHC class-I-dependent CD8⁺ T-cell response^{3–7}, a process referred to as cross-presentation². The cellular mechanisms allowing the cross-presentation pathway are poorly understood. Here we show that phagosomes display the elements and properties needed to be self-sufficient for the cross-presentation of exogenous antigens, a newly ascribed function linked to phagocytosis mediated by the endoplasmic reticulum.

Until recently, little was known about the molecular mechanisms and cellular structures enabling the cross-presentation pathway. A

requirement for peptide regurgitation from the phagosome lumen, allowing their direct binding to MHC class I molecules at the surface of presenting cells, was ruled out and a phagosome to cytosol pathway for the retrotranslocation of exogenous peptides was shown to exist⁸. Although peptide transport activity allowing for the retrotranslocation of exogenous molecules from the phagosome lumen to the cytoplasm has been observed in dendritic cells⁹, the phagosomal proteins involved in this process have not been identified. After peptide transfer to the cytosol, the proposed steps for cross-presentation involve: (1) ubiquitination and proteasomal degradation in the cytoplasm to generate the correct peptides for MHC class I loading, (2) peptide transport in the endoplasmic reticulum (ER) lumen through the TAP complex to form MHC class-I-peptide complexes, and (3) transport to the cell surface through the secretory pathway. The finding that phagocytosis in macrophages proceeds by ER recruitment at the cell surface, a process referred to as ER-mediated phagocytosis¹⁰, suggested that antigens from intracellular pathogens could have direct access, within phagosomes, to the ER machinery needed for MHC class I presentation.

Indications that phagosomes could be competent organelles for antigen cross-presentation came from proteomics analyses showing that proteins required for each step of the cross-presentation pathway, described above, are present on latex-bead-containing phagosomes (Supplementary Table 1). The cross-presentation of exogenous peptides requires a retrotranslocation step from the phagosome lumen to the cytoplasm to be processed by the proteasome. Confocal analysis showed that after internalization of latex beads opsonized with fluorescent ovalbumin, a fluorescent signal is observed in the cytoplasm near phagosomes (Fig. 1). Although the retrotranslocation machinery allowing this transfer is not known, a protein that might play this role on phagosomes is Sec61.

Indeed, Sec61 has been shown to mediate the retrotranslocation of proteins from the ER lumen to the cytoplasm^{11–12}, including the A1 subunit of cholera toxin (CTA1)¹³. When fluorescent CTA1 was examined in place of ovalbumin, a strong signal was also observed in the cytoplasm (Fig. 1), further suggesting that Sec61 might be involved in phagosomal retrotranslocation. In contrast to ovalbumin, which starts to be visible in the cytoplasm 60 min after phagocytosis of latex-OVA beads, the CTA1 signal is visible within 30 min. This difference might be explained by the fact that intra-phagosomal processing of ovalbumin is required before its retrotranslocation¹⁴. Interestingly, although the signal for CTA1 was also observed in the cytoplasm of cells not treated with MG-132, fluorescent ovalbumin was not observed in these conditions (results not shown). Because CTA1 is believed to be resistant to proteasomal degradation¹⁵, these results suggest that ovalbumin could be rapidly degraded by proteasomes close to or associated with phagosomes.

Further experiments performed to investigate the distribution of proteasomes in J774 murine macrophages showed that although proteasomes are present in the total cell lysate, indicative of a cytoplasmic localization, they are also observed in the total membrane preparation, largely made of ER in these cells, and on phagosomes where multiple proteasome α -subunits were detected (Fig. 2a). These results, in accordance with previous studies showing the presence of proteasomes in the cytoplasm, the nucleus and the ER^{16,17}, extend the cellular localization of this structure to phagosomes. Thus, the proteasomes present on phagosomes could come from the ER during ER-mediated phagocytosis. Analysis of the kinetics of association of proteasomes to phagosomes indicated that while Sec61 is recruited at the earliest time point, proteasome α -subunits associate transiently to phagosomes and reach a peak at around 60 min after their formation (Fig. 2b). This indicates that proteasomes do not come directly from the ER during phagosome formation, but rather assemble on phagosomes to play a function at a precise point during phagolysosome biogenesis.

Immunofluorescence analysis supported the western blot data by

showing a diffuse labelling for proteasome α - and β -subunits in the cytoplasm and around phagosomes (Fig. 2c). Immunofluorescence on isolated phagosomes¹⁰ indicated that about 70% of phagosomes displayed a discontinuous labelling for the proteasome α - and β -subunits (Fig. 2c, inset). Although this pattern of labelling is similar to the one previously observed for flotillin-1, a marker of lipid rafts on phagosomes¹⁸, proteasomes are not present in phagosome rafts (Fig. 2d). The proteasomes present on phagosomes are part of high-molecular-weight complexes, as shown by western blotting of purified phagosome preparations separated by non-denaturing blue native electrophoresis¹⁹ (Fig. 2e), in which the $\alpha 4$, $\alpha 6$, $\beta 1$, and $\beta 5$ subunits have been identified by mass spectrometry analyses. Finally, we were also able to show that proteasomes associate to the cytoplasmic side of phagosomes, based on their sensitivity to pronase (Fig. 2f).

Polyubiquitinated proteins are also present on the cytoplasmic side of phagosomes (pronase-sensitive) where they colocalize with proteasomes (Fig. 3a). A link between the ubiquitination process and proteasomal degradation on phagosomes was further highlighted when J774 macrophages were treated with the proteasome inhibitor MG-132. Indeed, the amount of polyubiquitinated proteins present on phagosomes increased with time when proteasome activity was inhibited (Fig. 3b), whereas it was lower and remained constant in untreated cells. These results are consistent with a continuous arrival and degradation of polyubiquitinated proteins on phagosomes. The nature of the polyubiquitinated proteins present on phagosomes was investigated using a proteomics approach, which allowed for the identification of 23 proteins, including ovalbumin (QTAMVLVNAIV and NVMEERK) present in at least five bands (Supplementary Table 2).

The improved ability of cells to process and present antigens on MHC class I molecules upon interferon- γ (IFN- γ) treatment is related to the upregulation of subsets of proteasomes, referred to as immunoproteasomes, differing from their 'steady state' counter-

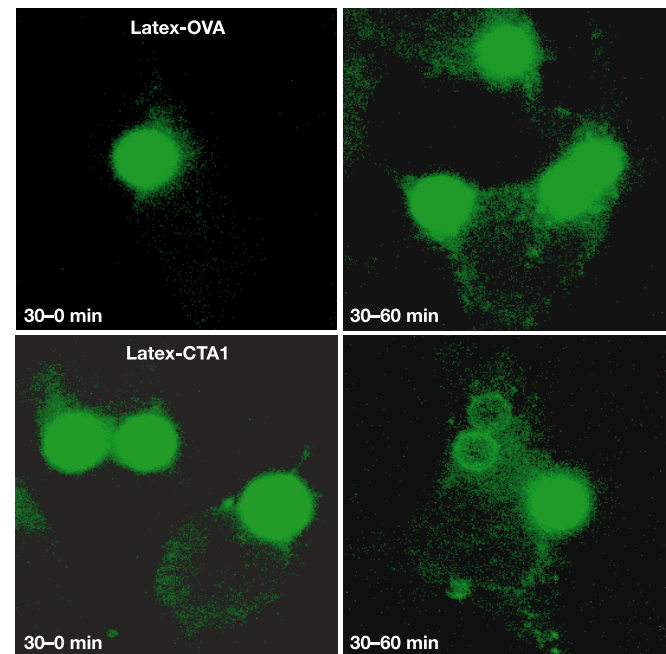


Figure 1 Exogenous proteins internalized by phagocytosis are retrotranslocated to the cytoplasmic side of phagosomes. For latex-OVA, after the initial 30 min of fluorescent ovalbumin-latex bead internalization, the fluorescent signal was contained within phagosomes (30 min–0 min). A fluorescent signal was observed in the cytoplasm near phagosomes after 60 min of chase (30 min–60 min). Latex-CTA1, the α -subunit of cholera toxin, is also translocated from the phagosome lumen, albeit more rapidly than ovalbumin.

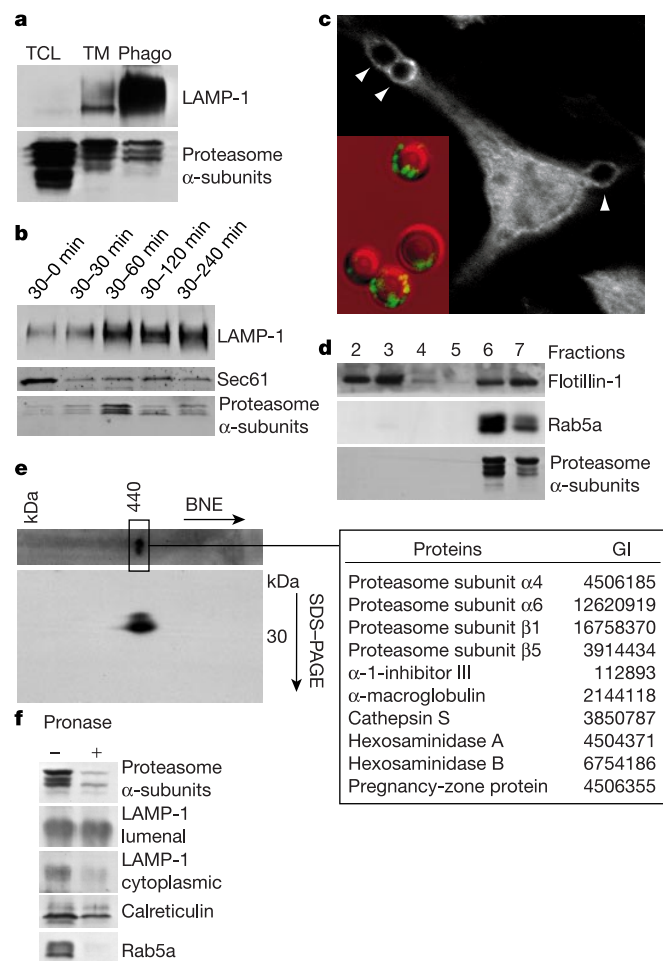


Figure 2 Proteasomes are present on phagosomes. **a**, Proteasome α-subunits are detected in the total cell lysate (TCL), the total membrane fraction (TM), and phagosomes (Phago). **b**, Proteasomes associate with phagosomes transiently, contrasting with the late association of LAMP1, and the early recruitment of Sec61. **c**, Immunofluorescence on cells and isolated organelles (inset) shows the association of proteasomes to phagosomes (arrowheads). **d**, Proteasomes are present in the TX-100 soluble fraction (non-raft) of phagosomes. Rab5a is a known non-raft protein. **e**, Blue native electrophoresis (BNE) indicates that proteasome α-subunits detected on phagosomes are present in a high-molecular-weight complex (440 kDa). Western blotting after SDS-PAGE confirmed the presence of α-subunits in that complex, while proteomics analysis allowed for the identification of several proteasome subunits. GI, GenBank identifier. **f**, Proteasomes are present on the cytoplasmic side of phagosomes as shown by their sensitivity to pronase, like rab5 and the cytoplasmic tail of LAMP1. Pronase does not affect the luminal part of LAMP1 and calreticulin, present in the phagosome lumen.

parts by the substitution of three β-subunits for the homologous subunits LMP2, LMP7 and MECL1. Treatment of J774 macrophages with IFN-γ resulted in the upregulation of the immunoproteasome subunit LMP2, as shown by immunofluorescence (Fig. 4a). Furthermore, western blot analysis indicated that immunoproteasomes associate with phagosomes upon IFN-γ treatment, a result confirmed by direct immunofluorescence on isolated organelles (Fig. 4a, inset). The peptides generated by proteasomal degradation could potentially gain access to the phagosome lumen since the TAP complex is present on this organelle, as shown by western blotting (Fig. 4b). Critical points that remained to be established were the presence of ovalbumin peptide/MHC class I complexes in the phagosome lumen, and how they are delivered to the cell surface for presentation to CD8⁺ T cells.

We showed that after phagocytosis of latex-OVA beads in BMA3.1A cells, the SIINFEKL ovalbumin peptide/MHC class I

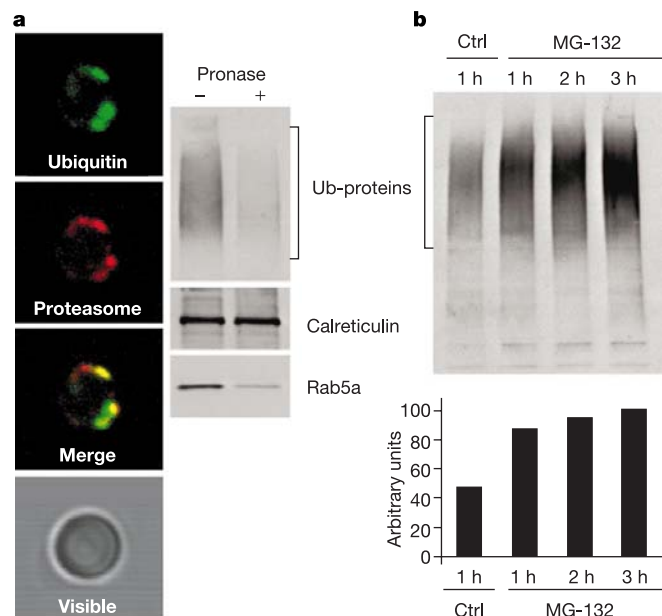


Figure 3 Ubiquitinated proteins associate with proteasomes on the cytoplasmic side of phagosomes. **a**, Ubiquitinated proteins colocalize with proteasomes on phagosomes (left panels show one phagosome). The pronase assay (right panels) indicates that polyubiquitinated proteins are present on the cytoplasmic side of phagosomes. **b**, The level of ubiquitinated proteins present on phagosomes isolated at different time points after formation remained constant (the 1-h time point is shown). In contrast, inhibition of proteasome activity with MG-132 led to the accumulation of polyubiquitinated proteins on phagosomes. Ctrl, control.

complex, recognized by the monoclonal antibody 25D1.16 (ref. 20), is present in the phagosome lumen (Fig. 4c), reaching a maximum level around 60–120 min after phagocytosis (not shown). Approximately 11% of the phagosomes contained detectable complexes in these conditions, an eightfold increase compared with phagosomes formed by the internalization of non-opsonized latex beads. The low signal here is due to the permeabilization procedure needed to detect molecules in the phagosome lumen. In these conditions, the value obtained for LAMP1 (around 20% in contrast to about 80–90% in non-permeabilized phagosome preparations) is the maximum expected value with that assay.

Bone-marrow-derived macrophages from B/6 mice that had internalized latex-OVA beads stimulated the proliferation of OT-1 CD8⁺ T cells and their secretion of IFN-γ (Fig. 4d, e), as shown previously⁸. Identical results were obtained with the C57BL/6-derived macrophage cell lines BMA3.1A and BMC2 (not shown). CD8⁺ T-cell proliferation and IFN-γ secretion were abolished when bone marrow-derived macrophages from TAP1^{-/-} mice were used, indicating the involvement of a transport event through TAP (Fig. 4d, e). Evidence that these events are triggered by MHC-peptide complexes originating from the phagosome lumen was obtained using a system measuring the percentage of OT-1 CD8⁺ T cells producing IFN-γ after a co-culture with macrophages infected with recombinant vaccinia viruses expressing NP-SIINFEKL-GFP²¹ or macrophages that had internalized latex-OVA beads (Fig. 4f).

The conventional presentation of the abundant SIINFEKL peptide/MHC class I complexes generated from endogenously derived ovalbumin peptides, synthesized in the cell cytoplasm following vaccinia infection and reaching the surface through the secretory pathway, was almost completely inhibited by brefeldin A (Fig. 4f). In contrast, the presentation of complexes generated after phagocytosis of latex-OVA beads was only partially (35%) inhibited by this drug (Fig. 4f). The IFN-γ response was significantly greater than the

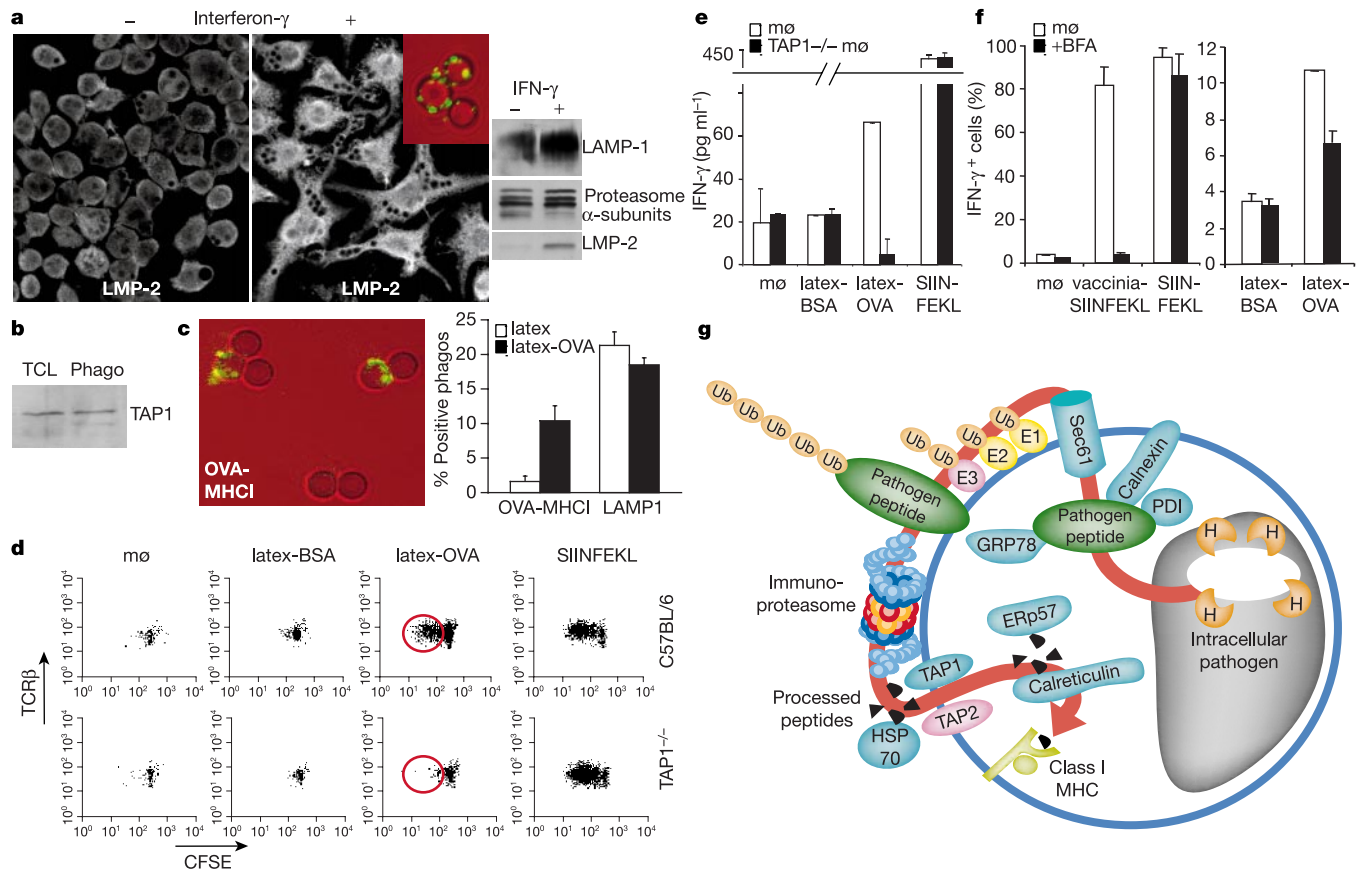


Figure 4 Exogenous proteins loaded in phagosomes can be presented by MHC class I complexes at the surface of macrophages and trigger a TAP-dependent CD8⁺ T-cell response. **a**, Treatment with IFN- γ increases the level of proteasome LAMP-2 subunits in J774 macrophages and their association to phagosomes. **b**, TAP is present on phagosomes. **c**, The SIINFEKL/MHC I complex is present in the phagosome lumen. **d**, Internalization of latex-OVA in C57BL/6 macrophages triggers the proliferation of OT-I transgenic CD8⁺ T-cells (red circle). Macrophages without bead (m ϕ), or that had internalized latex-BSA did not induce the proliferation of T cells. As a positive control, incubation of macrophages with SIINFEKL peptide, allowing direct binding to MHC I at the cell surface, sustained the proliferation of T cells. The ability of macrophages to elicit a

CD8 response is TAP-dependent as shown by the absence of CD8⁺ T-cell proliferation after latex-OVA internalization by macrophages from TAP1 $^{-/-}$ mice. **e**, The same trend was observed for the secretion of IFN- γ by CD8⁺ T cells cultured with the various macrophages. **f**, A significant level of presentation of SIINFEKL/MHC I complexes following phagocytosis of latex-OVA is observed in cells treated with brefeldin A. In contrast, activation of T cells was almost completely abolished when macrophages infected with recombinant vaccinia viruses expressing NP-SIINFEKL-green fluorescent protein (GFP) were treated with BFA. BFA had no effect on the activation of T cells by the direct addition of the SIINFEKL peptide. **g**, Working model. See text.

background response elicited by macrophages following uptake of latex-BSA beads. The cell surface delivery of MHC-peptide complexes from endo/phagocytic organelles was also shown to occur via a brefeldin-A-insensitive route^{3,22}, so the significant level of cross-presentation of ovalbumin observed in the brefeldin-A-treated cells further supports the phagosome origin of the peptide/MHC class I complexes involved in CD8⁺ T-cell activation.

Our results demonstrate that phagosomes are able to process exogenous peptides for MHC class I cross-presentation, extending the competence of phagosomes previously shown to be functional for the processing of MHC class II complexes²³. Our working model (Fig. 4g) proposes that hydrolases (H) acquired sequentially during phagosome maturation initiate the processing of exogenous peptides in the phagosome lumen^{24–26}. Some of these peptides are then retrotranslocated to the cytoplasmic side of phagosomes by the Sec61/chaperones retrotranslocation machinery normally used for quality control^{11,27}. The translocated proteins then have access to the ubiquitin/proteasome complex assembled on the cytoplasmic side of phagosomes, leading to the generation of MHC class I binding peptides. Although some of these peptides are likely to be transported in the ER lumen, some reach the lumen of phagosomes through the TAP complex present on phagosomes. The membrane recycling machinery of endo/phagocytic organelles²⁸ could then be used for

the delivery of MHC class-I-peptide complexes to the surface.

The constant sampling of self and foreign molecules by the immune system is part of a complex quality control process ensuring the recognition and clearance of aberrant cell forms during development, tumorigenesis and infection by microbial pathogens. Our results, and those presented in an accompanying paper²⁹, establish that the use of ER, an organelle specialized in quality control, to form part of the phagosome membrane serves not only to minimize the utilization of the plasma membrane, but also confers properties allowing phagosomes to be fully integrated within the immune recognition system and play a direct role in cross-presentation. Our results, together with findings showing that ovalbumin linked to beads internalized by phagocytosis is presented with MHC class I molecules up to 1×10^4 -fold more efficiently than soluble ovalbumin³⁰, suggest that favouring the entry of exogenous peptides by phagocytosis could improve immunization procedures for vaccination protocols and the development of more efficient immunotherapies against cancer. □

Methods

Antibodies

Mouse anti-proteasome α -subunits monoclonal antibodies (mAb), rabbit anti-proteasome α/β -subunit polyclonal antibodies (pAb), rabbit anti-LAMP-2 pAb and mouse

anti-ubiquitin clone FK1 mAb were from Affiniti. Rabbit anti-rab5a pAb and goat anti-TAP1 mAb were from Santa Cruz. Rat anti-LAMP1 luminal ID4B mAb was from Developmental Studies Hybridoma Bank. Mouse anti-CD8 (CyChrome-conjugated), mouse anti-IFN- γ (PE-conjugated), and hamster anti-TCR (PE-conjugated) were from BD PharMingen. Rabbit anti-LAMP1 cytoplasmic pAb was a gift from S. M  resse. Rabbit anti-flotillin pAb was a gift from R. Parton. Rabbit anti-Sec61p pAb was a gift from C. Nicchitta. Rabbit anti-calreticulin pAb was a gift from L. Rokeach. Mouse anti-OVA-MHCI complex 25D1.16 mAb was a gift from R. Germain and J. Yewdell.

Phagosome formation and isolation

All cell types were cultured as described^{7,8,10}. Phagosomes were formed by the internalization of 0.8- μ m latex beads and isolated as described¹⁰. Phagosome lipid rafts were isolated after Triton X-100 solubilization and separation on an Optiprep gradient, as described¹⁸.

Retrotranslocation assay

To determine if exogenous proteins present within the phagosome lumen can be retrotranslocated, phagosomes were formed in J774 macrophages by a 30 min internalization of latex beads opsonized with ovalbumin or the α -subunit of cholera toxin labelled with Alexa Fluor 488 according to standard Molecular Probes protocol. After an additional 60 min of incubation in the presence of 5 μ M MG-132, an inhibitor of proteasome activity, cells were fixed and prepared for confocal microscopy.

Immunofluorescence on cells and isolated phagosomes

Immunofluorescence on cells was performed using standard procedures described previously¹⁰. For the immunoproteasome, J774 cells were treated with IFN- γ 36 h before phagosome formation. For all direct immunofluorescence analyses, isolated phagosomes containing 3.0- μ m latex beads were fixed in paraformaldehyde 4% for 10 min then washed three times with PBS/BSA 3% by quick spins. For the detection of the SIINFEKL/class I complex in the phagosome lumen, permeabilization was required. In this case, after fixation, phagosomes were treated with 1% Triton X-100 at room temperature for 10 min and then washed three times with PBS/BSA 3%. Phagosomes were then blocked for 30 min with PBS/BSA 3% and the primary antibody added for 45 min. Phagosomes were washed three times and fluorescent secondary antibody was added for 30 min. After three washes, phagosomes were prepared for the confocal microscope.

Blue native electrophoresis

To determine whether the proteasome subunits detected by fluorescence on phagosomes are present as a complex, we isolated phagosomes and performed blue native gel electrophoresis following a protocol described previously¹⁹.

T-cell proliferation

CD8⁺ T cells from OT-I CD8⁺ transgenic mice were negatively selected by magnetic separation (MACS system and CD8⁺ T-cell purification kit, Miltenyi Biotec) according to the manufacturer's indications. Purified CD8⁺ T cells were further labelled with the intracellular fluorescent dye carboxyfluorescein diacetate succinimidyl ester (CFSE; Molecular Probes). The cells were resuspended at 5×10^7 cells ml⁻¹ in PBS with 0.5 μ M CFSE for 10 min at 37 $^{\circ}$ C, and the reaction was stopped with 10% normal mouse serum. Cells were washed with cold PBS 0.1% BSA and plated at 10^6 per well in a 24-well plate in DMEM 10% FCS. Antigen-pulsed macrophages were added for 72 h, at which time the cells were fixed in 4% paraformaldehyde. T-cell proliferation was measured by flow cytometry as expressed by the intensity of CFSE labelling. The lymphocytes were identified by characteristic size and granularity, in combination with anti-TCR β chain and anti-CD8 surface staining. For each sample, a minimum of 20,000 events was collected and analysed using CellQuest software and a FACSCalibur flow cytometer (BD Biosciences).

Brefeldin A treatment and interferon secretion assays

To assess the transport of MHC-peptide complexes to the cell surface, BMA3.1A macrophages were incubated with latex-OVA or vaccinia-OVA²¹ with or without brefeldin A (10 μ g ml⁻¹) for 6–10 h, at which time CD8⁺ T cells were added for an additional 4 h with BFA. For detection of intracellular IFN- γ , CD8⁺ T cells were further fixed, permeabilized and stained following standard procedures⁷. IFN- γ production in 72 h cell culture supernatants was quantified by ELISA using Endogen's antibody matched pairs and following the manufacturer's protocol.

Proteomics analysis

Phagosome proteins were separated by SDS-PAGE on 5-cm minigels. Gel slices (24 in total) were excised, trypsin digested, and the resulting tryptic peptides extracted with 0.2 M urea in 50% aqueous acetonitrile. Each digested band was analysed by nanoLC-MS/MS using a Waters CapLC coupled to Q-TOF Ultima. A total of 460 proteins were identified, including several proteins involved in each step of cross-presentation (Supplementary Table 1). The nature of the polyubiquitinated proteins present on phagosomes was also investigated using proteomics. In this case, after phagosome isolation, the polyubiquitinated proteins present on this organelle were purified by affinity for the S5A subunit of the proteasome attached to Sepharose beads (Affiniti) (Supplementary Table 2). These proteins were then separated and analysed as above.

Received 14 March; accepted 15 July 2003; doi:10.1038/nature01912.

1. Watts, C. & Amigorena, S. Phagocytosis and antigen presentation. *Semin. Immunol.* **13**, 373–379 (2001).

- Heath, W. R. & Carbone, F. R. Cross-presentation in viral immunity and self-tolerance. *Nature Rev. Immunol.* **1**, 126–134 (2001).
- Pfeifer, J. D. *et al.* Phagocytic processing of bacterial antigens for class I MHC presentation to T cells. *Nature* **361**, 359–362 (1993).
- Oliveira, S. C. & Splitter, G. A. CD8⁺ type 1 CD44hi CD45 RBlo T lymphocytes control intracellular *Brucella abortus* infection as demonstrated in major histocompatibility complex class I- and class II-deficient mice. *Eur. J. Immunol.* **25**, 2551–2557 (1995).
- Turner, J. & Dockrell, H. M. Stimulation of human peripheral blood mononuclear cells with live *Mycobacterium bovis* BCG activates cytolytic CD8⁺ T cells *in vitro*. *Immunology* **87**, 339–342 (1996).
- Canaday, D. H. *et al.* Activation of human CD8⁺ α beta TCR⁺ cells by *Mycobacterium tuberculosis* via an alternate class I MHC antigen-processing pathway. *J. Immunol.* **162**, 372–379 (1999).
- Belkaid, Y. *et al.* CD8⁺ T cells are required for primary immunity in C57BL/6 mice following low-dose, intradermal challenge with *Leishmania major*. *J. Immunol.* **168**, 3992–4000 (2002).
- Kovacsovics-Bankowski, M. & Rock, K. L. A phagosome-to-cytosol pathway for exogenous antigens presented on MHC class I molecules. *Science* **267**, 243–246 (1995).
- Rodriguez, A., Regnault, A., Kleijmeer, M., Ricciardi-Castagnoli, P. & Amigorena, S. Selective transport of internalized antigens to the cytosol for MHC class I presentation in dendritic cells. *Nature Cell Biol.* **1**, 362–368 (1999).
- Gagnon, E. *et al.* Endoplasmic reticulum-mediated phagocytosis is a mechanism of entry into macrophages. *Cell* **110**, 119–131 (2002).
- Wiertz, E. J. *et al.* Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction. *Nature* **384**, 432–438 (1996).
- Tirosh, B., Furman, M. H., Tortorella, D. & Ploegh, H. L. Protein unfolding is not a prerequisite for ER-to-cytosol dislocation. *J. Biol. Chem.* **278**, 6664–6672 (2003).
- Schmitz, A., Herrgen, H., Winkler, A. & Herzog, V. Cholera toxin is exported from microsomes by the Sec61p complex. *J. Cell Biol.* **148**, 1203–1212 (2000).
- Rodriguez, A., Regnault, A., Kleijmeer, M., Ricciardi-Castagnoli, P. & Amigorena, S. Selective transport of internalized antigens to the cytosol for MHC class I presentation in dendritic cells. *Nature Cell Biol.* **1**, 362–368 (1999).
- Hazes, B. & Read, R. J. Accumulating evidence suggests that several AB-toxins subvert the endoplasmic reticulum-associated protein degradation pathway to enter target cells. *Biochemistry* **36**, 11051–11054 (1997).
- Rivett, A. J., Palmer, A. & Knecht, E. Electron microscopic localization of the multicatalytic proteinase complex in rat liver and in cultured cells. *J. Histochem. Cytochem.* **40**, 1165–1172 (1992).
- Brooks, P., Murray, R. Z., Mason, G. G., Hendil, K. B. & Rivett, A. J. Association of immunoproteasomes with the endoplasmic reticulum. *Biochem. J.* **352**, 611–615 (2000).
- Dermine, J. F. *et al.* Flotillin-1-enriched lipid raft domains accumulate on maturing phagosomes. *J. Biol. Chem.* **276**, 18507–18512 (2001).
- Schagger, H., Cramer, W. A. & von Jagow, G. Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal. Biochem.* **217**, 220–230 (1994).
- Porgador, A., Yewdell, J. W., Deng, Y., Binnink, J. R. & Germain, R. N. Localization, quantitation, and *in situ* detection of specific peptide-MHC class I complexes using a monoclonal antibody. *Immunity* **6**, 715–726 (1997).
- Princiotta, M. F. *et al.* Quantitating protein synthesis, degradation, and endogenous antigen processing. *Immunity* **18**, 343–354 (2003).
- MacAry, P. A. *et al.* Mobilization of MHC class I molecules from late endosomes to the cell surface following activation of CD34-derived human Langerhans cells. *Proc. Natl Acad. Sci. USA* **98**, 3982–3987 (2001).
- Ramachandra, L., Song, R. & Harding, C. V. Phagosomes are fully competent antigen-processing organelles that mediate the formation of peptide: class II MHC complexes. *J. Immunol.* **162**, 3263–3272 (1999).
- Garin, J. *et al.* The phagosome proteome: insight into phagosome functions. *J. Cell Biol.* **152**, 165–180 (2001).
- Muno, D., Kominami, E. & Mizuuchi, T. Generation of both MHC class I- and class II-restricted antigenic peptides from exogenously added ovalbumin in murine phagosomes. *FEBS Lett.* **478**, 178–182 (2000).
- Lennon-Dumenil, A. M. *et al.* Analysis of protease activity in live antigen-presenting cells shows regulation of the phagosomal proteolytic contents during dendritic cell activation. *J. Exp. Med.* **196**, 529–540 (2002).
- Tsai, B., Ye, Y. & Rapoport, T. A. Retro-translocation of proteins from the endoplasmic reticulum into the cytosol. *Nature Rev. Mol. Cell Biol.* **3**, 246–255 (2002).
- Mahnke, K. *et al.* The dendritic cell receptor for endocytosis, DEC-205, can recycle and enhance antigen presentation via major histocompatibility complex class II-positive lysosomal compartments. *J. Cell Biol.* **151**, 673–684 (2000).
- Guermonez, P. *et al.* ER-phagosome fusion defines an MHC class I cross-presentation compartment in dendritic cells. *Nature* **425**, 397–402 (2003).
- Kovacsovics-Bankowski, M., Clark, K., Benacerraf, B. & Rock, K. L. Efficient major histocompatibility complex class I presentation of exogenous antigen upon phagocytosis by macrophages. *Proc. Natl Acad. Sci. USA* **90**, 4942–4946 (1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank C. Rondeau, S. Tessier and A. Carrier for technical assistance and G. Milon for discussions. We thank K. Rock for the BMA3.1A and BMC2 cell lines. We also thank D. Boismenu and J. Bergeron from the Montreal Proteomics Network for help with the mass spectrometry analyses of the blue native gel samples. This work was supported by the Canadian Institute for Health Research, Genome Canada/Qu  bec (M.D.), and the NIH (D.S.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.D. (michel.desjardins@umontreal.ca).

A high-affinity conformation of Hsp90 confers tumour selectivity on Hsp90 inhibitors

Adeela Kamal, Lia Thao, John Sensintaffar, Lin Zhang, Marcus F. Boehm, Lawrence C. Fritz & Francis J. Burrows

Conforma Therapeutics Corporation, 9393 Towne Centre Drive, Suite 240, San Diego, California 92121, USA

Heat shock protein 90 (Hsp90) is a molecular chaperone that plays a key role in the conformational maturation of oncogenic signalling proteins, including HER-2/ErbB2, Akt, Raf-1, Bcr-Abl and mutated p53¹⁻⁷. Hsp90 inhibitors bind to Hsp90, and induce the proteasomal degradation of Hsp90 client proteins^{6,8-11}. Although Hsp90 is highly expressed in most cells, Hsp90 inhibitors selectively kill cancer cells compared to normal cells, and the Hsp90 inhibitor 17-allylaminogeldanamycin (17-AAG) is currently in phase I clinical trials^{12,13}. However, the molecular basis of the tumour selectivity of Hsp90 inhibitors is unknown. Here we report that Hsp90 derived from tumour cells has a 100-fold higher binding affinity for 17-AAG than does Hsp90 from normal cells. Tumour Hsp90 is present entirely in multi-chaperone complexes with high ATPase activity, whereas Hsp90 from normal tissues is in a latent, uncomplexed state. *In vitro* reconstitution of chaperone complexes with Hsp90 resulted in increased binding affinity to 17-AAG, and increased ATPase activity. These results suggest that tumour cells contain Hsp90

complexes in an activated, high-affinity conformation that facilitates malignant progression, and that may represent a unique target for cancer therapeutics.

The naturally occurring ansamycin antibiotic geldanamycin (GM) binds to a conserved binding pocket in the amino-terminal domain of Hsp90^{14,15}, inhibiting ATP binding and ATP-dependent Hsp90 chaperone activity¹⁶⁻¹⁸. The GM derivative 17-AAG exerts anti-tumour activity in preclinical models¹⁹ and is currently in clinical trials^{12,20}. However, two paradoxical observations have confounded understanding of the preferential effects of Hsp90 inhibitors on tumour cells. First, it is unclear why 17-AAG binds purified Hsp90 protein with micromolar affinity *in vitro*²¹ but has nanomolar activity in tissue culture¹. Second, although Hsp90 is abundantly expressed in most cells, ansamycin drugs selectively accumulate in human tumour xenografts *in vivo*^{12,22} (our unpublished observations). It has been speculated that the physico-chemical properties of 17-AAG may contribute to cellular accumulation²³, but the mechanism is unknown¹². Previous work showing that the binding affinity of GM to Hsp90 from cell lysates¹ was greater than that for the isolated protein²¹ suggested to us that ansamycin binding to Hsp90 could be influenced by additional intracellular factors. Furthermore, if the unidentified factors were present to greater extent in tumour cells than normal cells, differential binding affinity might also explain the tumour selectivity of Hsp90 inhibitors.

To investigate if Hsp90 from tumour cells had a higher binding affinity to 17-AAG than that from normal cells, we performed competitive binding assays using a biotinylated GM (biotin-GM) probe (Fig. 1a). Incubation of Hsp90 from BT474 breast carcinoma cells with increasing concentrations of 17-AAG inhibited the binding of Hsp90 to biotin-GM with an IC₅₀ of 6 nM, whereas Hsp90

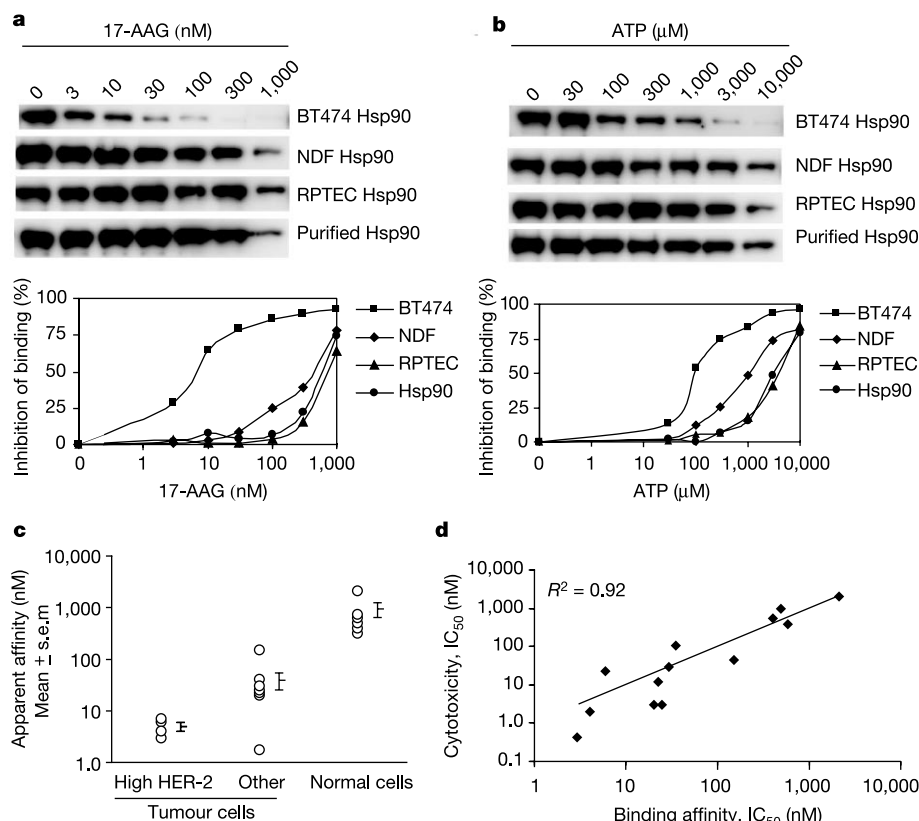


Figure 1 17-AAG has a higher binding affinity to Hsp90 from tumour cells than normal cells or purified Hsp90 protein. **a**, **b**, 17-AAG (**a**) or ATP (**b**) has a higher binding affinity to Hsp90 from BT474 tumour cells than normal cells (NDF and RPTEC) and purified Hsp90. **c**, Apparent binding affinity of 17-AAG to Hsp90 from a panel of tumour and

normal cells. Individual values of IC₅₀ (circles) and the mean ± standard error of the mean (line with error bars) is shown. **d**, Correlation of the binding affinity of 17-AAG to Hsp90 to the cytotoxicity in different cells.

from primary human renal epithelial cells (RPTEC) and normal dermal fibroblasts (NDF) had an IC_{50} of 600 nM and 400 nM, respectively. Purified Hsp90 protein binding was inhibited with an IC_{50} of 600 nM, as previously reported²¹. The apparent binding affinity of ATP, the physiological ligand of Hsp90, was also higher in the tumour cell lines ($IC_{50} = 100 \mu M$ in BT474) than the normal cells ($IC_{50} = 1,000 \mu M$ in NDF and $2,900 \mu M$ in RPTEC) (Fig. 1b), whereas purified Hsp90 had an IC_{50} of $3,000 \mu M$. In addition, by testing a large panel of cell lysates (see Methods), we observed that Hsp90 from HER-2-overexpressing cancer cells (BT474, N87, SKOV3 and SKBR3) had IC_{50} values of 5 ± 1 nM, that from other tumour cells (MCF7, A549, HT29, MDA468, SKMG3, U87, HT1080 and Hs578t) was 39 ± 14 nM, whereas Hsp90 from normal human primary cells (NDF, RPTEC, HMVEC, HMEC, HUVEC, Hs578Bst and PBMC) had IC_{50} values of 943 ± 290 nM (Fig. 1c). Total Hsp90 protein levels did not vary greatly between tumour and normal cells (Supplementary Fig. 1a). Taken together, these results suggest that the Hsp90 in tumour cells, particularly those overexpressing the Hsp90 client protein HER-2, has a significantly higher binding affinity for 17-AAG than does Hsp90 from normal cells. Furthermore, the binding affinity of 17-AAG to Hsp90 from the different cells directly correlates with the cytotoxic/cytostatic activity of 17-AAG in those cells (Fig. 1d).

Hsp90 interacts with many co-chaperone proteins that assemble into multi-chaperone complexes that promote protein folding²⁴. Tumour cells overexpress Hsp90 client proteins, suggesting that a

greater amount of Hsp90 in tumour cells might be engaged in active chaperoning and present in multi-chaperone complexes that could modulate the binding affinity of ligands to Hsp90. To assess whether Hsp90 in tumour cells was present in multi-chaperone complexes, we determined the levels of the co-chaperones p23 and Hop—essential components of the two known multi-chaperone Hsp90 complexes¹³—associated with Hsp90 from normal and tumour cells (Fig. 2a). Co-immunoprecipitation in the normal and tumour cell lysates with antibodies to Hsp90 revealed that more Hsp90 in tumour cells was present in complexes with p23 and Hop compared to normal cells (Fig. 2a). Control immunoprecipitations without antibodies did not immunoprecipitate any Hsp90. Strikingly, immunoprecipitation with antibodies to both p23 and Hop revealed that the entire tumour cell pool of Hsp90 was present in complexes unlike normal cells. Conversely, immunoprecipitation with an antibody²⁵ specific for the uncomplexed form of Hsp90 pulled down far more Hsp90 from normal cells than from tumour cells. Neither the total levels of p23 and Hop nor the growth rate of the tumour cells differed significantly from the normal cells (Supplementary Fig. 1b). These results suggest that all Hsp90 in tumour cells is in the form of multi-chaperone complexes that are actively engaged in chaperoning client oncoproteins, and that the heightened complex formation is not due to increased expression of Hsp90 or co-chaperones. As the chaperone function of Hsp90 is dependent upon its ATPase activity¹⁶, we immunoprecipitated Hsp90 from cell lysates and performed ATPase assays. Tumour Hsp90 had markedly higher ATPase activity compared to Hsp90 from normal cells (Fig. 2b) and was inhibited by 17-AAG, radicicol (another structurally unrelated Hsp90 inhibitor), and by 0.5 M KCl (which is known to break apart chaperone complexes). Control immunoprecipitation in the absence of Hsp90 antibody did not have any significant ATPase activity (data not shown). Combined, these results suggest that essentially all soluble Hsp90 in tumour

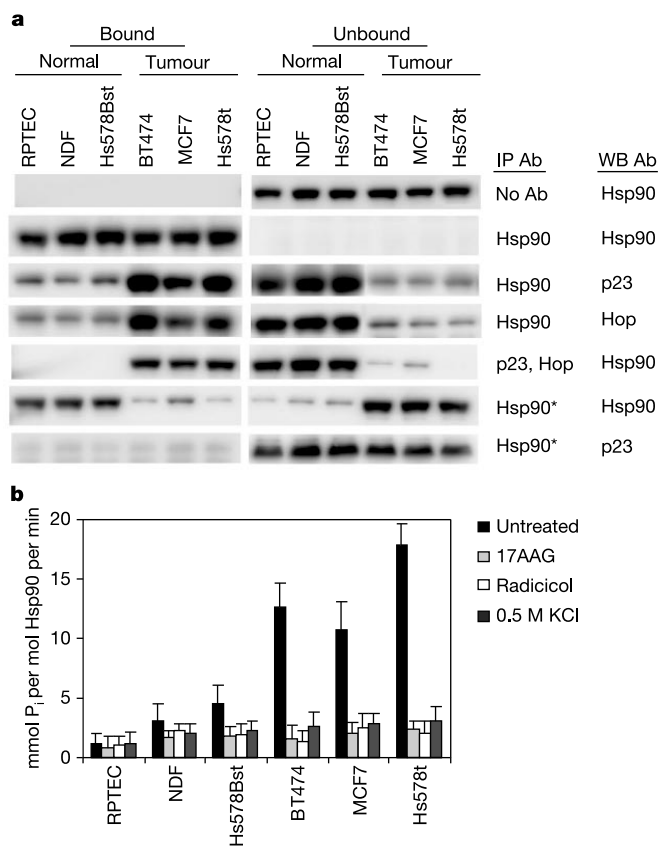


Figure 2 Hsp90 from tumour cells is present in chaperone complexes with high ATPase activity. **a**, Normal (RPTEC, NDF and Hs578Bst) and tumour (BT474, MCF7 and Hs578t) cell lysates were immunoprecipitated with indicated antibodies (IP Ab) and immunoblotted with indicated western blot (WB) Ab. The bound and unbound fractions are shown. Hsp90* antibody preferentially recognizes uncomplexed Hsp90. **b**, ATPase activity of Hsp90 from normal (RPTEC, NDF and Hs578Bst) and tumour (BT474, MCF7 and Hs578t) cell lysates that were untreated or treated with 10 μM 17-AAG, 10 μM radicicol and 0.5 M KCl. Mean $\pm$ standard deviation are shown from triplicate experiments.

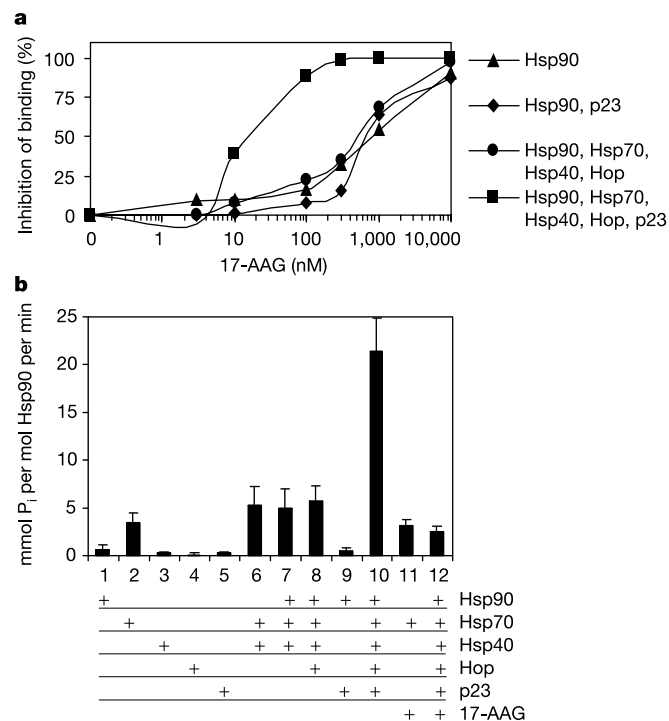


Figure 3 *In vitro* reconstitution of purified Hsp90 with co-chaperones increased binding affinity to 17-AAG and the ATPase activity. **a**, Hsp90 reconstituted with the indicated co-chaperones has a higher apparent binding affinity to 17-AAG. **b**, Hsp90 reconstituted with the indicated co-chaperones has increased ATPase activity, which can be inhibited by 10 μM 17-AAG.

cells is present in fully active multi-chaperone complexes, whereas Hsp90 in normal cells is in an uncomplexed inactive form.

As our data demonstrated that the Hsp90 from tumour cells had a higher binding affinity to 17-AAG, and this correlates with presence of increased multi-chaperone complexes and increased ATPase activity, we tested whether reconstitution of purified Hsp90 with co-chaperones would result in increased binding affinity and ATPase activity. For *in vitro* reconstitution, we used five proteins (Hsp90, Hsp70, Hsp40, Hop and p23) that have been shown to be required for the *in vitro* chaperoning activity of Hsp90²⁶. Addition of all four co-chaperones to purified Hsp90 increased the apparent affinity of 17-AAG from 600 nM for Hsp90 alone to 12 nM in the presence of the added proteins, whereas partial complexes did not (Fig. 3a). Therefore, Hsp90 when reconstituted with co-chaperones has nanomolar binding activity, in concordance with that observed in tumour cell lysates (Fig. 1). The ATPase activity of Hsp90 was also significantly enhanced by all four co-chaperones and was inhibited by the addition of 17-AAG (Fig. 3b). Hsp70 had some minimal ATPase activity, but this was not inhibited by 17-AAG. Thus, these results suggest that Hsp90 present in multi-chaperone complexes not only had a higher binding affinity for 17-AAG but was also more biochemically active.

To determine if our observations *in vitro* also applied to mice and to clinical cancer, we examined the binding affinity of 17-AAG to Hsp90 in normal and malignant mouse and human tissue samples (Fig. 4a, b). The apparent binding affinity of 17-AAG to Hsp90 from mouse tumours (3T3-src, B16 and CT26) was 8–35 nM compared to 200–600 nM for the mouse normal tissues (brain, kidney, liver, lung and heart), even though Hsp90 was more abundantly expressed in several of the normal tissues (Fig. 4a). For the human tissues, using four samples of each tissue type, the IC₅₀ was 6,170 ± 1,060 nM for normal breast versus 29 ± 4 nM for

breast carcinoma, and 2,725 ± 736 nM for normal colon versus 43 ± 4 nM for the colon carcinomas (Fig. 4b). To determine the fraction of Hsp90 present as multi-chaperone complexes, we performed co-immunoprecipitations and observed that there were increased amounts of p23 with Hsp90 from the human tumours compared to the normal tissues (Fig. 4c). Conversely, there was significantly more uncomplexed Hsp90 co-immunoprecipitated with the Hsp90* antibody from the normal tissues compared to tumour tissues (Fig. 4c). Furthermore, the Hsp90 ATPase activity in both clinical tumour samples was significantly higher relative to the normal tissues and was inhibited by 17-AAG, radicicol and 0.5M KCl (Fig. 4d). These results indicate that the Hsp90 in human clinical tumour tissues is in a high-affinity, multi-chaperone complex with increased ATPase activity.

Our data show that Hsp90 in tumour cells exists in a functionally distinct molecular form, and clarify three fundamental aspects of the role of Hsp90 in tumour biology. First, the nanomolar binding of 17-AAG to high-affinity tumour Hsp90 is now consistent with the nanomolar anti-tumour activity of this family of drugs. Second, the markedly higher affinity of tumour Hsp90 *in vivo* explains the remarkable ability of ansamycins to accumulate progressively at tumour sites in animals. And third, the complete usage of tumour Hsp90 may provide a selection pressure leading to further upregulation of Hsp90 that is observed in many advanced tumours^{27,28}. We propose a model of Hsp90-dependent malignant progression in which, as tumour cells gradually accumulate mutant and over-expressed signalling proteins, Hsp90 becomes engaged in active chaperoning and stabilization of oncoproteins, and adopts a high-affinity form induced by bound co-chaperone proteins. Thus, Hsp90 may serve an analogous role in the somatic evolution of tumours as in darwinian evolution^{29,30}—rescuing potentially misfolded mutant proteins to prevent the mutation becoming lethal to

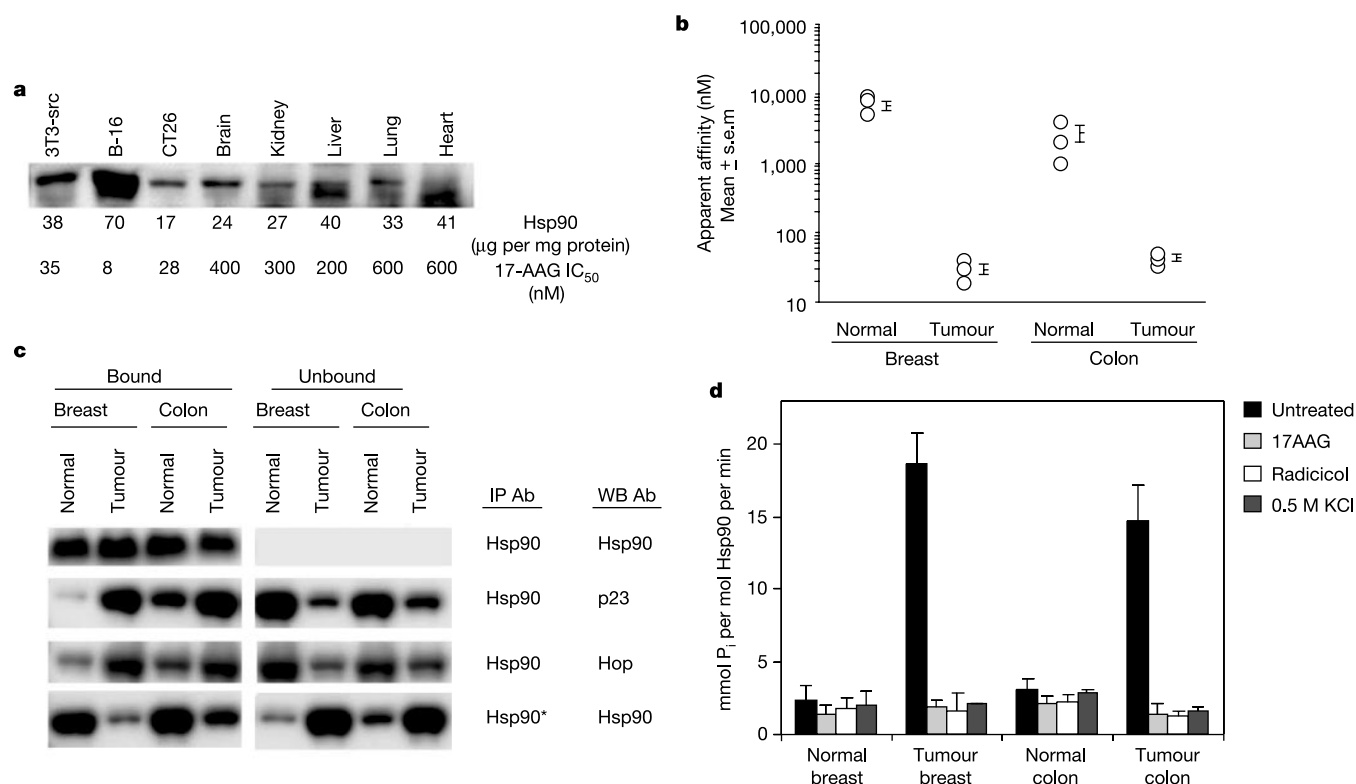


Figure 4 Hsp90 from clinical tumour samples is in a high-affinity complex with increased ATPase activity. **a**, Hsp90 protein levels and apparent binding affinity of 17-AAG to Hsp90 from mouse tumour and normal tissues. **b**, Apparent binding affinity of 17-AAG to Hsp90 from human normal and tumour tissues. Mean ± standard error is shown ($N = 4$).

c, Human normal and tumour tissue lysates co-immunoprecipitated with Hsp90 antibodies and immunoblotted with indicated WB Ab. **d**, Hsp90 ATPase activity from human normal and tumour tissues treated with 10 μM 17-AAG or radicicol and 0.5 M KCl. Mean ± standard error is shown ($N = 4$).

the cell, and thus allowing these oncoproteins to support tumour cell proliferation, survival and malignancy. Dependence on the activated, high-affinity chaperone could make Hsp90 an 'Achilles heel' of tumour cells, driving the selective accumulation and bioactivity of pharmacological Hsp90 inhibitors, and making tumour Hsp90 a unique cancer target. □

Methods

Cells and reagents

Normal and malignant cell lines were obtained from Clonetics and ATCC unless otherwise indicated. Tumour cells used included high-HER-2 breast carcinoma (BT474, SKBR3), stomach carcinoma (N87) and ovarian carcinoma (SKOV3), MCF7 breast carcinoma, MDA468 breast carcinoma, HS578T breast carcinoma and paired normal epithelium HS578Bst, A549 lung carcinoma, HT29 colon carcinoma, U87 glioblastoma, SKMG3 glioblastoma (a gift from C. Thomas) and HT1080 fibrosarcoma. Normal primary cells used were renal proximal epithelial cells (RPTEC), normal dermal fibroblasts (NDF), human microvascular endothelial cells (HMVEC), human mammary epithelial cells (HMEC), human umbilical vascular endothelial cells (HUVCE) and peripheral blood mononuclear cells (PBMC). Normal mouse tissue (brain, kidney, liver, lung and heart) was obtained from nude mice and the mouse tumours were from nude mice injected with 3T3-src (a gift from L. Neckers), B16 and CT26 cells. The clinical samples (benign breast tissue, metastatic ductal breast carcinoma, normal colonic mucosa and colon adenocarcinoma) were obtained from Bainbridge Genomic Foundation and BioResearch Support, Inc. Antibodies used were: Hsp90 (Stressgen, SPA-845; recognizes Hsp90 α and Hsp90 β and immunoprecipitates free and complexed Hsp90), Hsp90* (Stressgen, SPA-830; recognizes Hsp90 α and Hsp90 β and immunoprecipitates uncomplexed Hsp90), p23 (Alexis Biochemicals, 804-023-R100) and Hop (a gift from D. Toft). 17-AAG was synthesized from GM, and the biotin-GM probe was prepared by displacing the 17-methoxy of GM with a biotinyl-linked amine.

Hsp90 binding assays

Purified native Hsp90 protein (Stressgen) or cell lysates in lysis buffer (20 mM HEPES, pH 7.3, 1 mM EDTA, 5 mM MgCl₂, 100 mM KCl) were incubated with or without 17-AAG for 30 min at 4 °C, and then incubated with biotin-GM linked to BioMag streptavidin magnetic beads (Qiagen) for 1 h at 4 °C. Tubes were placed on a magnetic rack, and the unbound supernatant removed. The magnetic beads were washed three times in lysis buffer and heated for 5 min at 95 °C in SDS-PAGE sample buffer. Samples were analysed on SDS protein gels, and western blots done using indicated antibodies. Bands in the western blots were quantified using the Bio-rad Fluor-S Multimager, and the percentage inhibition of binding of Hsp90 to the biotin-GM was calculated. The IC₅₀ reported is the concentration of 17-AAG needed to cause half-maximal inhibition of binding. For *in vitro* reconstitution, 5 μ M of purified Hsp90 (Stressgen) was combined with 1 μ M each of Hsp70, Hsp40, p23 and Hop purified proteins (a gift from D. Toft) as previously described²⁶.

MTS assay

Cells were seeded in 96-well plates at 2,000 cells per well in a final culture volume of 100 μ l for 24 h before the addition of increasing concentrations of 17-AAG that was incubated for 5 days. Viable cell number was determined using the Celltiter 96 Aqueous Non-radioactive Cell Proliferation Assay (Promega). The value of the background absorbance at 490 nm (A₄₉₀) of wells not containing cells was subtracted. Percentage of viable cells = (A₄₉₀ of 17-AAG treated sample/A₄₉₀ untreated cells) $\times$ 100. The IC₅₀ was defined as the concentration that gave rise to 50% viable cell number.

Co-immunoprecipitation

Cell lysates were prepared as described above. Protein-A Sepharose beads (Zymed) were pre-blocked with 5% BSA. The cell lysates were pre-cleared by incubating with 50 μ l of protein-A Sepharose beads (50% slurry). To 100 μ l of the pre-cleared cell lysate, either no antibody or antibodies to Hsp90, p23 and Hop were added, and incubated by rotating for 1 h at 4 °C. 50 μ l of pre-cleared beads (50% slurry) was then added and incubated by rotating for 1 h at 4 °C. Bound beads were briefly centrifuged at 3,000g and unbound samples collected. Beads were washed thrice in lysis buffer and once with 50 mM Tris, pH 6.8, and then SDS-sample buffer added for 5 min at 95 °C. Bound and unbound samples were analysed by SDS-PAGE and western blots using indicated antibodies.

Hsp90 ATPase assay

Hsp90 was co-immunoprecipitated from normal and tumour cell lysates as described above using an N-terminal Hsp90 antibody. For *in vitro* reconstitution, 5 μ M Hsp90 was combined with 1 μ M each of Hsp70, Hsp40, p23 and Hop purified proteins²⁵. The ATPase activity of the immunoprecipitated or *in vitro* reconstituted Hsp90 was measured by detection of free inorganic phosphate (P_i) using a PiPer Phosphate Assay kit (Molecular Probes). The assay measures an increase in fluorescence absorption of an Amplex Red reagent that is proportional to the amount of P_i in the sample. Standards were done using known concentrations of P_i (100 μ M with six half-log dilutions), and the amount of P_i released from Hsp90 in the cell lysates or in the *in vitro* reconstituted proteins was calculated. For the cell lysates, the amount of Hsp90 immunoprecipitated with the Hsp90 antibody was quantified to normalize the data to mmol Pi per mol Hsp90 per min. All assays were done in triplicate.

Received 4 June; accepted 11 July 2003; doi:10.1038/nature01913.

1. Schulte, T. W. & Neckers, L. M. The benzoquinone ansamycin 17-allylamino-17-

demethoxygeldanamycin binds to HSP90 and shares important biologic activities with geldanamycin. *Cancer Chemother. Pharmacol.* **42**, 273–279 (1998).

- Xu, W. *et al.* Sensitivity of mature ErbB2 to geldanamycin is conferred by its kinase domain and is mediated by the chaperone protein Hsp90. *J. Biol. Chem.* **276**, 3702–3708 (2001).
- Sato, S., Fujita, N. & Tsuruo, T. Modulation of Akt kinase activity by binding to Hsp90. *Proc. Natl Acad. Sci. USA* **97**, 10832–10837 (2000).
- Basso, A. D., Solit, D. B., Munster, P. N. & Rosen, N. Ansamycin antibiotics inhibit Akt activation and cyclin D expression in breast cancer cells that overexpress HER2. *Oncogene* **21**, 1159–1166 (2002).
- Schulte, T. W., Blagosklonny, M. V., Ingui, C. & Neckers, L. Disruption of the Raf-1-Hsp90 molecular complex results in destabilization of Raf-1 and loss of Raf-1-Ras association. *J. Biol. Chem.* **270**, 24585–24588 (1995).
- An, W. G., Schulte, T. W. & Neckers, L. M. The heat shock protein 90 antagonist geldanamycin alters chaperone association with p210bcr-abl and v-src proteins before their degradation by the proteasome. *Cell Growth Differ.* **11**, 355–360 (2000).
- Blagosklonny, M. V., Toretsky, J. & Neckers, L. Geldanamycin selectively destabilizes and conformationally alters mutated p53. *Oncogene* **11**, 933–939 (1995).
- Whitesell, L., Minnaugh, E. G., De Costa, B., Myers, C. E. & Neckers, L. M. Inhibition of heat shock protein HSP90-p60v-src heteroprotein complex formation by benzoquinone ansamycins: Essential role for stress proteins in oncogenic transformation. *Proc. Natl Acad. Sci. USA* **91**, 8324–8328 (1994).
- Miller, P. *et al.* Depletion of the erbB-2 gene product p185 by benzoquinoid ansamycins. *Cancer Res.* **54**, 2724–2730 (1994).
- Minnaugh, E. G., Chavany, C. & Neckers, L. Polyubiquitination and proteasomal degradation of the p185c-erbB-2 receptor protein-tyrosine kinase induced by geldanamycin. *J. Biol. Chem.* **271**, 22796–22801 (1996).
- Schulte, T. W., An, W. G. & Neckers, L. M. Geldanamycin-induced destabilization of Raf-1 involves the proteasome. *Biochem. Biophys. Res. Commun.* **239**, 655–659 (1997).
- Workman, P. Auditing the pharmacological accounts for hsp90 molecular chaperone inhibitors: Unfolding the relationship between pharmacokinetics and pharmacodynamics. *Mol. Cancer Ther.* **2**, 131–138 (2003).
- Isaacs, J. S., Xu, W. & Neckers, L. Heat shock protein 90 as a molecular target for cancer therapeutics. *Cancer Cell* **3**, 213–217 (2003).
- Prodromou, C. *et al.* Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. *Cell* **90**, 65–75 (1997).
- Stebbins, C. E. *et al.* Crystal structure of an Hsp90-geldanamycin complex: Targeting of a protein chaperone by an antitumor agent. *Cell* **89**, 239–250 (1997).
- Obermann, W. M., Sondermann, H., Russo, A. A., Pavletich, N. P. & Hartl, F. U. In vivo function of Hsp90 is dependent on ATP binding and ATP hydrolysis. *J. Cell Biol.* **143**, 901–910 (1998).
- Panaretou, B. *et al.* ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo. *EMBO J.* **17**, 4829–4836 (1998).
- Grenert, J. P., Johnson, B. D. & Toft, D. O. The importance of ATP binding and hydrolysis by hsp90 in formation and function of protein heterocomplexes. *J. Biol. Chem.* **274**, 17525–17533 (1999).
- Solit, D. B. *et al.* 17-Allylamino-17-demethoxygeldanamycin induces the degradation of androgen receptor and HER-2/neu and inhibits the growth of prostate cancer xenografts. *Clin. Cancer Res.* **8**, 986–993 (2002).
- Blagosklonny, M. V. Hsp-90-associated oncoproteins: Multiple targets of geldanamycin and its analogs. *Leukemia* **16**, 455–462 (2002).
- Chiosis, G. *et al.* A small molecule designed to bind to the adenine nucleotide pocket of Hsp90 causes Her2 degradation and the growth arrest and differentiation of breast cancer cells. *Chem. Biol.* **8**, 289–299 (2001).
- Egorin, M. J. *et al.* PC3 human prostate xenograft retention of, and oncoprotein modulation by 17-(allylamino)-17-demethoxygeldanamycin(17AAG) in vivo. *Proc. Am. Assoc. Cancer Res.* **40**, 3409 (1999).
- Chiosis, G. *et al.* 17AAG: Low target binding affinity and potent cell activity-finding an explanation. *Mol. Cancer Ther.* **2**, 123–129 (2003).
- Richter, K. & Buchner, J. Hsp90: Chaperoning signal transduction. *J. Cell. Physiol.* **188**, 281–290 (2001).
- Schuh, S. *et al.* A 90,000-dalton binding protein common to both steroid receptors and the Rous sarcoma virus transforming protein, pp60v-src. *J. Biol. Chem.* **260**, 14292–14296 (1985).
- Kosano, H., Stensgaard, B., Charlesworth, M. C., McMahon, N. & Toft, D. The assembly of progesterone receptor-hsp90 complexes using purified proteins. *J. Biol. Chem.* **273**, 32973–32979 (1998).
- Gress, T. M. *et al.* Differential expression of heat shock proteins in pancreatic carcinoma. *Cancer Res.* **54**, 547–551 (1994).
- Yano, M., Naito, Z., Tanaka, S. & Asano, G. Expression and roles of heat shock proteins in human breast cancer. *Jpn J. Cancer Res.* **87**, 908–915 (1996).
- Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).
- Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417**, 618–624 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Toft for critical reading of the manuscript. We also thank G. Timony for preparing the scatter plots, P. Karjian for quantitative western blotting of Hop and p23, and other team members at Conformia Therapeutics Corporation for discussions and suggestions.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to F.J.B. (fburrows@conformacorp.com).

A micrococcal nuclease homologue in RNAi effector complexes

Amy A. Caudy^{1*}, René F. Ketting^{2*}, Scott M. Hammond^{1†}, Ahmet M. Denli¹, Anja M. P. Bathoorn^{1,2}, Bastiaan B. J. Tops^{1,2}, Jose M. Silva¹, Mike M. Myers¹, Gregory J. Hannon¹ & Ronald H. A. Plasterk²

¹Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA

²The Hubrecht Laboratory, Centre for Biomedical Genetics, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

* These authors contributed equally to this work

† Present address: Department of Cell and Developmental Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA

RNA interference (RNAi) regulates gene expression by the cleavage of messenger RNA, by mRNA degradation and by preventing protein synthesis. These effects are mediated by a ribonucleoprotein complex known as RISC (RNA-induced silencing complex)¹. We have previously identified four *Drosophila* components (short interfering RNAs¹, Argonaute 2 (ref. 2), VIG and FXR³) of a RISC enzyme that degrades specific mRNAs in response to a double-stranded-RNA trigger. Here we show that Tudor-SN (tudor staphylococcal nuclease)—a protein containing five staphylococcal/micrococcal nuclease domains and a tudor domain—is a component of the RISC enzyme in *Caenorhabditis elegans*, *Drosophila* and mammals. Although Tudor-SN contains non-canonical active-site sequences, we show that purified Tudor-SN exhibits nuclease activity similar to that of other staphylococcal nucleases. Notably, both purified Tudor-SN and RISC are inhibited by a specific competitive inhibitor of micrococcal nuclease. Tudor-SN is the first RISC subunit to be identified that contains a recognizable nuclease domain, and could therefore contribute to the RNA degradation observed in RNAi.

Exposure of cells to double-stranded RNA (dsRNA) can elicit various types of sequence-specific gene silencing⁴. A signature of these silencing events is the involvement of small RNAs of approximately 22–25 nucleotides (nt) that guide the selection of silencing targets^{1,5,6}. These short interfering RNAs (siRNAs) or microRNAs (miRNAs) are generated by the processing of silencing triggers by an RNaseIII family nuclease, Dicer⁷. Small RNAs join multicomponent ribonucleoprotein (RNP) complexes, known generically as RISCs, which enforce silencing.

Both to address the nature of the RNAi effector machinery in detail, and to examine the relationship between the different effector mechanisms of RNAi, we have biochemically purified a RISC complex from *Drosophila* that degrades its mRNA target, and have sought to identify its protein and RNA components. In multiple, independent purifications of RISC, we identified, along with previously characterized proteins, a potentially novel component corresponding to a *Drosophila* candidate gene, CG7008 (Supplementary Fig. 1). This evolutionarily conserved 103 kDa protein contains five repeats of a staphylococcal/micrococcal nuclease domain (Supplementary Fig. 2). Four of these repeats are intact, whereas the fifth repeat is fused at its amino terminus to a tudor domain, which has been implicated in the binding of modified amino acids⁸. On the basis of this characteristic domain structure, we named the protein Tudor-SN, for tudor staphylococcal nuclease. Through each purification step, Tudor-SN co-fractionated with known RISC components (Fig. 1a, b and Supplementary Fig. 3).

Orthologues of Tudor-SN are found in plants (*Arabidopsis*⁹), *C. elegans*^{9,10}, mammals^{10,11} and *Schizosaccharomyces pombe* (A.M.D., unpublished observations), but not in *Saccharomyces*

cerevisiae. To investigate whether a role for Tudor-SN orthologues in RNAi is evolutionarily conserved, we carried out biochemical fractionation of extracts from *C. elegans* and mammalian cells.

We began by preparing cytosolic extracts from synchronized cultures of wild-type *C. elegans*. As in *Drosophila*, a large fraction of the miRNA population can be extracted from the ribosomes (Supplementary Fig. 4). Size fractionation of extracts derived from adult animals revealed that miRNAs eluted from the column in two peaks, representing ~500 kDa and ~250 kDa complexes (Fig. 1c), similar to what had been observed previously in extracts from *Drosophila* S2 cells³. Three different miRNAs—*lin-4*, *let-7* and *mir-52*—behaved identically in this assay. By contrast, size fractionation of *C. elegans* egg extract, and examination of complexes containing *mir-40* and *mir-52*, suggested the presence of only the 500 kDa complex (Fig. 1c). Thus, it seems that miRNAs in *C. elegans* can inhabit multiple, distinct RNP complexes, and that the partitioning of miRNAs between these complexes may depend on both the identity of the miRNA and the developmental stage of the organism.

RISC complexes in *C. elegans* have not previously been characterized. We therefore probed whether *Drosophila* RISC components co-fractionated with miRNAs in *C. elegans* extracts. We raised antibodies to F56D12.5 (VIG-1), the worm homologue of *Drosophila* VIG, and F10G7.2 (TSN-1), the worm orthologue of Tudor-SN. Both VIG-1 and TSN-1 were enriched in the fractions that contained 250 kDa miRNA (Fig. 1c). By contrast, VIG-1 or TSN-1 did not appear in fractions containing 500 kDa miRNA complex.

To test whether putative RISC components are present in

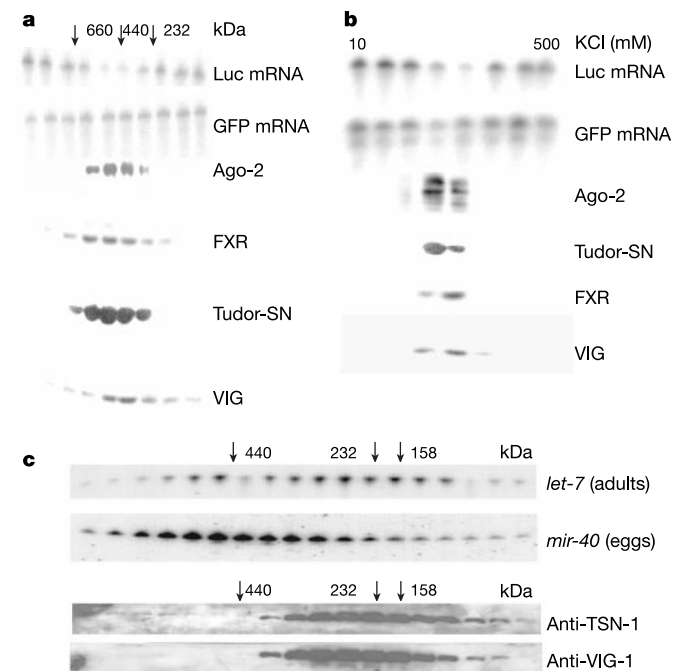


Figure 1 Identification and confirmation of Tudor-SN as a component of RISC complexes. **a**, *Drosophila* RISC activity was extracted from ribosomes and fractionated on Superose 6. The migration of size markers is indicated. RISC assays and controls for nonspecific activity were carried out as in ref. 1. Western blots were done using antibodies to Ago-2, VIG, FXR and Tudor-SN. **b**, The active gel-filtration fractions from **a** were further chromatographed on Source Q. Fractions were analysed as in **a**. **c**, Size fractionation of *C. elegans* ribosome-associated extracts prepared from both eggs and adults. The fractions were analysed by northern blotting. Egg-derived extract is probed for *mir-40*, the adult extract for *let-7*. Similar results are obtained when these blots are probed for *mir-52*. Below, size-fractionated extract from adult animals, analysed by western blotting as indicated for TSN-1 and VIG-1. GFP, green fluorescent protein; Luc, luciferase.

the same complex, we used antibodies to immunoprecipitate individual components. In *Drosophila*, antibodies to either FXR or VIG co-immunoprecipitated Argonaute 2 (Ago-2; Fig. 2a), as was predicted by our previous findings using epitope-tagged versions of these proteins³. Similar amounts of Ago-2 were also co-immunoprecipitated using antibodies directed against Tudor-SN (Fig. 2a). Furthermore, antibodies directed against FXR and VIG co-immunoprecipitate Tudor-SN (Fig. 2a), and VIG and Tudor-SN antisera conversely recover FXR (Fig. 2a), indicating that all four proteins are present in a single complex. In *C. elegans*, VIG-1-specific antibodies can co-immunoprecipitate TSN-1 (Fig. 2b), indicating a similar association of the *C. elegans* orthologues of *Drosophila* RISC proteins.

In naive mammalian cells, we had difficulty detecting interactions between orthologues of *Drosophila* RISC components (Fig. 2c). However, the formation of a complex containing these proteins was

induced if we first triggered an RNAi response by transfection with siRNAs. Specifically, we showed interactions between an Argonaute family protein, AGO2 (human GERP95/EIF2C2/AGO2; reviewed in ref. 12), the fragile X mental retardation protein (FMRP) and the mammalian Tudor-SN homologue, p100 (Fig. 2c). Notably, complex formation occurred without changes in the expression levels of individual RISC components, indicating that association of pre-existing proteins is nucleated when a siRNA becomes available in the cell (Fig. 2c).

RISC is an RNP complex that can contain either siRNAs or miRNAs^{1–3,13}. Consistent with their roles as components of RISC, both miRNAs (Fig. 2d) and siRNAs (Fig. 2e) can be co-immunoprecipitated from *Drosophila* cells using antisera that recognize FXR, VIG and Tudor-SN. Similarly, in *C. elegans*, TSN-1 and VIG-1 immunoprecipitates contain *let-7* RNA (Fig. 2f). In addition, we find the mammalian *let-7* miRNA in immunoprecipitates of p100, and in parallel confirm the previously demonstrated association between miRNAs and AGO2 (refs 13–15; Fig. 2g). We also detect miRNAs in association with members of the fragile X family in mammalian cells, including FMRP, FXR1 and FXR2 (S.M.H., unpublished observations). Considered together, our results point to a common architecture for RISC in animals as a complex that contains a small RNA (miRNA or siRNA) and protein components that include an Argonaute family member, VIG, Tudor-SN and, at least in *Drosophila* and mammals, a fragile X family member.

The mammalian homologue of Tudor-SN, known as p100, has

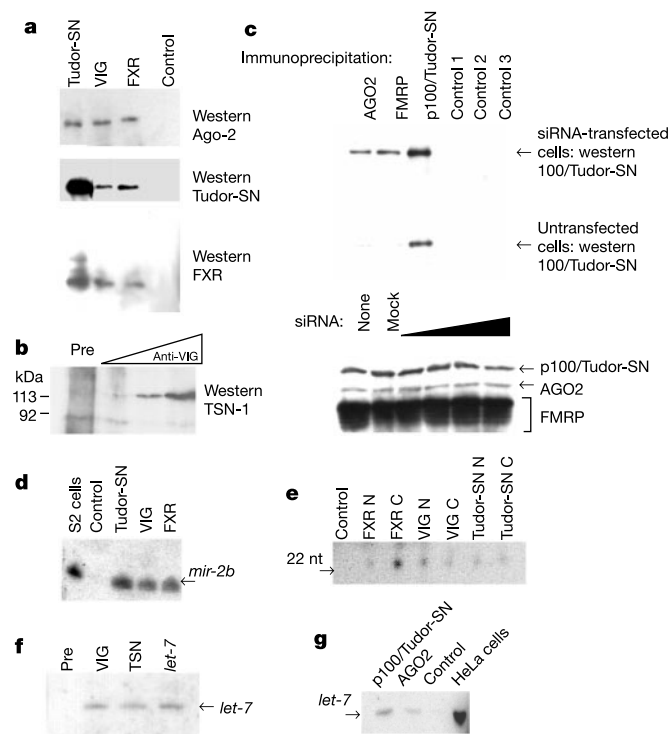


Figure 2 Immunoprecipitations from multiple organisms confirm association between Tudor-SN and components of RISC. **a**, Immunoprecipitations from *Drosophila* S2 cells were done using the indicated affinity-purified antibody or a control antibody—affinity-purified antibody to the Drosha protein. Western blots of the immunoprecipitates were as indicated. **b**, Immunoprecipitations from *C. elegans* were performed with a nonspecific serum (Pre) or an increasing amount of VIG-1-specific antibody (2, 5 and 10 μ l anti-VIG). Precipitates were western-blotted using an anti-TSN-1 antibody. **c**, Immunoprecipitations from 293 cells were done using the indicated antibodies, and precipitates were blotted for p100/Tudor-SN. The control antibodies were as follows: control 1, non-immunized rabbit serum; control 2, irrelevant mouse ascites; control 3, non-immunized guinea pig serum. As indicated, the top panel of immunoprecipitations was done using cells previously transfected with siRNAs corresponding to luciferase, the bottom panel was done with untransfected cells. Below, westerns as indicated on cells untransfected (none), mock-transfected (mock) or transfected with siRNAs at a concentration of 10–100 nM. **d**, RNA was prepared from the indicated *Drosophila* immunoprecipitates as in **a**, and northern blotted for *mir-2b*, a miRNA expressed in S2 cells. **e**, *Drosophila* S2 cells were transfected with radiolabelled dsRNA. Two days after transfection, immunoprecipitations were performed as in **a**, and RNA was analysed on a 15% denaturing polyacrylamide gel. **f**, Immunoprecipitation from *C. elegans* was performed with a nonspecific serum (Pre), VIG-1-specific antibodies or TSN-1-specific antibodies. RNA was isolated from immunoprecipitated material and blotted for *let-7*. **g**, RNA was prepared from the indicated HeLa cell immunoprecipitates and blotted for *let-7*.

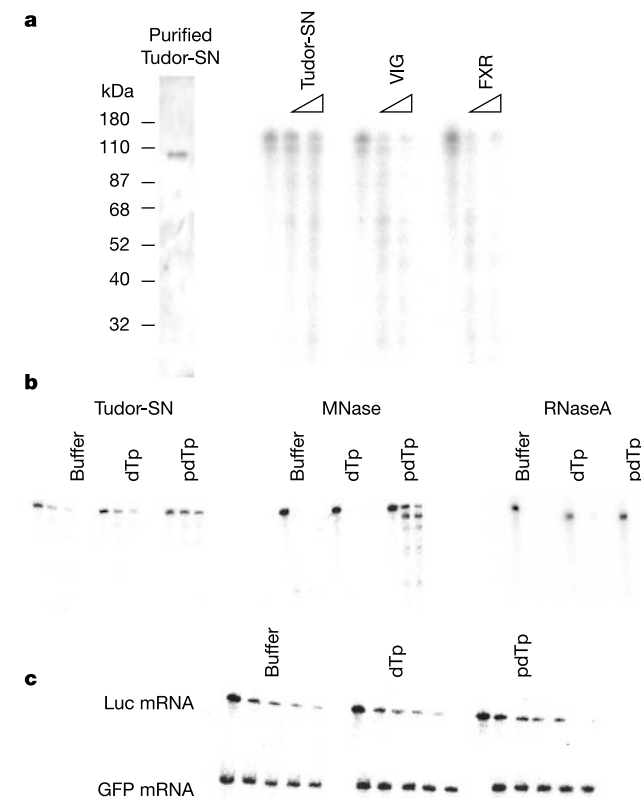


Figure 3 Tudor-SN has nuclease activity. **a**, 6 $\times$ His N-tagged *Drosophila* Tudor-SN was purified to homogeneity. A silver-stained gel is shown. Purified Tudor-SN fractions were immunodepleted with the indicated antibodies. The immunodepleted supernatants were tested for activity against single-stranded RNA. **b**, Purified Tudor-SN was treated with 100 μ M pdTp, dTp or a buffer control, substrate RNAs were added, and reaction time points of 0, 30 and 60 min were taken. **c**, Fractions containing partially purified RISC were treated for 15 min with 100 μ M pdTp, dTp or a buffer control, substrate RNAs were added, and reaction time points of 0, 15, 30, 45 and 60 min were taken.

been implicated as a co-activator for an Epstein–Barr virus transcription factor, EBNA-2 (ref. 10). An exclusively nuclear localization of Tudor-SN would be inconsistent with a role in RNAi, as many studies of RNAi in *C. elegans*, *Drosophila*, *Neurospora* and mammals have shown that post-transcriptional gene silencing by RNAi occurs largely in the cytoplasm. In both *Drosophila* and mammalian cells, Tudor-SN/p100 was present predominantly in cytoplasmic fractions (Supplementary Figs 5, 6). Examination of Tudor-SN immunoreactivity also showed a predominantly cytoplasmic localization (Supplementary Fig. 7; J.M.S., unpublished observations; ref. 16). In *C. elegans*, TSN-1 is found in significant amounts both in the nucleus and in the cytosol (not shown).

RISC is a nuclease that catalyses endonucleolytic cleavage of substrates, as directed by the associated siRNA. In many cases, siRNAs also trigger mRNA degradation, and the results of our biochemical purification are consistent with an association between RISC and a nuclease that catalyses complete destruction of targeted mRNAs¹. To assess the possibility that Tudor-SN might contribute to catalysis by RISC, we examined its intrinsic nuclease activity. All five staphylococcal nuclease domains of Tudor-SN contain mutations that alter the canonical active site, as derived from comparisons of bacterial family members and from structural data^{9,10}. A large panel of mutations has been made in staphylococcal nuclease, some of which are similar to those that alter Tudor-SN domains away from the active-site consensus¹⁷. Generally, these mutations lower the reaction rate but do not abolish catalysis.

We expressed and purified recombinant *Drosophila* Tudor-SN

from *Escherichia coli* (Fig. 3a). Nuclease activity precisely co-fractionated with the recombinant protein (not shown), and mono-specific carboxy-terminal Tudor-SN anti-peptide antibodies selectively depleted activity from purified Tudor-SN preparations (Fig. 3b). In these depletion experiments, nuclease activity was recovered on antibody–sepharose complexes (not shown). Micrococcal nucleases show broad substrate specificity, cutting both RNA and DNA. Similarly, Tudor-SN can cleave both RNA and DNA substrates (not shown).

3',5'-Deoxythymidine bisphosphate (pdTp), known to be a specific competitive inhibitor of staphylococcal nucleases¹⁸, inhibits Tudor-SN at 100 μ M concentrations, whereas dTp (3'-deoxythymidine monophosphate) does not inhibit Tudor-SN activity (Fig. 3d). Importantly, RISC activity is also inhibited by pdTp (Fig. 3e). These data are consistent with the possibility that Tudor-SN contributes at least some of the nuclease activity observed in RNAi effector complexes, but are also consistent with other interpretations (see below).

To examine the role of Tudor-SN in dsRNA-mediated silencing *in vivo*, we made use of a reporter system in *C. elegans*. A *lacZ*–*lin-41* fusion transgene expresses LacZ in the seam cells under the translational control of *let-7* (ref. 19). In wild-type animals, LacZ staining in the seam cells can be observed from the L1 stage through to L4. The staining is absent in adult animals as a consequence of *let-7*-mediated repression (Fig. 4a, b). As expected, suppression of *alg-1* and *dcr-1* by RNAi resulted in persistence of LacZ staining in adults^{20–22} (Fig. 4c, d). Next, we used RNAi to analyse the involvement of VIG-1 and TSN-1 in *let-7* function. RNAi against either VIG-1 or TSN-1 results in persistent expression of the LacZ reporter in the seam cells of adult animals (Fig. 4e, f). This shows that VIG-1 and TSN-1 are required for proper function of the *let-7* miRNA *in vivo*. By contrast, after RNAi against VIG-1, TSN-1 and DCR-1, we do not see an effect on RNAi efficiency. The effects we observe after RNAi against VIG-1, TSN-1 and DCR-1 (refs 20–22) probably reflect a partial reduction of function, as other phenotypes associated with *let-7* loss of function (vulva and alae defects²³) are not observed.

Our data strongly indicate that Tudor-SN is a bona fide RISC component. This is reflected by the co-purification of Tudor-SN and RISC in *Drosophila*, *C. elegans* and mammalian cells. However, it remains open to question whether Tudor-SN is a catalytic engine of RNAi. Despite the aforementioned data, there are potential inconsistencies. First, purified, recombinant Tudor-SN is non-sequence-specific, in contrast to RISC, which shows a high degree of selectivity for its mRNA targets. Second, Tudor-SN will cleave both RNA and DNA, whereas we detect no DNase activity in RISC (data not shown). Third, several investigators have detected specific cleavage of mRNAs within the siRNA–mRNA hybrid, and this is difficult to rationalize with the known activities of Tudor-SN and related enzymes^{6,13,24,25}. It is certainly consistent with our biochemical data to suppose that RISC contains multiple nucleases, only one of which (the putative Slicer) can catalyse site-specific mRNA cleavage. In this scenario, Tudor-SN might act to degrade the remainder of the mRNA. In accord with this idea, targeting of an mRNA by a single siRNA often results in complete degradation of the mRNA^{26,27}. Alternatively, it is possible that Tudor-SN does not have a catalytic role in the RISC complex. Indeed, pdTp is a competitive inhibitor that engages the potential nucleic-acid-binding domains of Tudor-SN. Thus, inhibition of RISC by pdTp may reflect a block in the ability of Tudor-SN to engage RNAs, possibly including the mRNA target, in the context of the RISC complex. Answers to these questions will come only from understanding RISC in sufficient detail to allow reconstitution of its native activity from purified components such that we can study in detail the individual contributions of each to the varied roles of the RNAi effector machinery. □

Methods

Drosophila cell culture and extract preparation

S2 cells were soaked with luciferase dsRNA at 3 mg l^{−1} and harvested after 7 days. Hypotonic

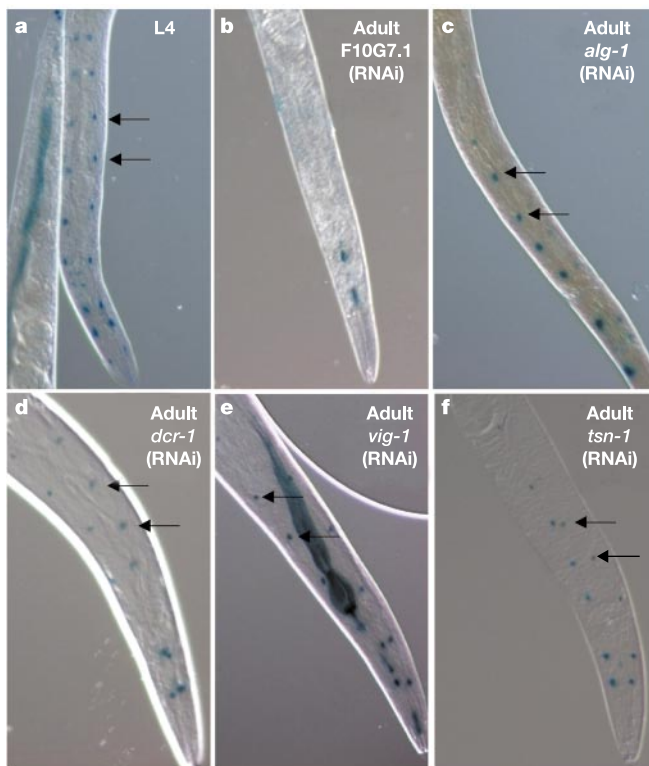


Figure 4 Staining of animals carrying a transgene expressing LacZ in the seam cells, under the translational control of *let-7*. In L4-stage animals, LacZ is expressed (a), but in adult animals the LacZ expression is repressed, also after performing RNAi against a control gene that is in the same operon as TSN-1 (22 of 23 animals silenced) (b). RNAi against genes that are involved in miRNA function and processing (*alg-1* and *dcr-1*) results in the re-expression of LacZ in adult animals (c, 24 of 28 animals; d, 15 of 29 animals). RNAi against *vig-1* and *tsn-1* also results in re-expression of LacZ in adults (e, 17 of 33 animals; f, 14 of 32 animals).

extracts were spun at 200,000g for 3 h to pellet ribosomes, which were resuspended and extracted in 400 mM potassium acetate. Resultant soluble protein was precipitated by 1:4 dilution in hypotonic buffer and redissolved in buffer A (20 mM HEPES (pH 7.0), 2 mM MgCl₂, 2 mM dithiothreitol and 0.5% octyl glucoside) with 400 mM KCl. This protein was then fractionated on two Superose-6 HR10/10 columns (Pharmacia) in series. Fractions were assayed for RISC activity as described¹. The fraction showing peak activity was diluted 1:5 in buffer A and fractionated over a Mono Q HR5/5 column (Pharmacia) with a linear gradient elution of buffer A with 0–500 mM KCl. The RISC fraction with peak activity from the Mono Q column was used for the EDTA/EGTA experiments described here. The purification of RISC for Tudor-SN identification was essentially as described².

C. elegans culture and extract preparation

Nematodes were grown in liquid culture. At the appropriate stage, the animals were collected by sedimentation on ice, and purified by flotation on 30% sucrose. After washing, the animals were resuspended in 0.5 volumes of hypotonic buffer: 10 mM HEPES (pH 7.1), 5 mM MgCl₂, 2 mM DTT and protease inhibitors (Complete protease inhibitor tablets; Roche). The suspension was dropped into liquid nitrogen, and the resulting balls were ground in a mortar. The powder was allowed to thaw on ice, and was centrifuged for 10 min at 14,000g. The supernatant was used as crude extract. The crude extract was then spun at 117,000g for 2 h to obtain a non-ribosome-associated supernatant fraction (S100) and a ribosome-pellet fraction (P100). The P100 was resuspended in hypotonic buffer supplemented with 500 mM NaCl, and mixed at 4 °C for 2 h. This was again separated in a ribosome-pellet fraction and a ribosome-associated supernatant fraction (S100H) by centrifugation at 117,000g for 2 h. The S100H fraction was used in most experiments. For egg extracts, animals were grown on bacteria that expressed *pos-1* dsRNA. Eggs were purified from sucrose-purified gravid adults by bleach treatment, and the resulting egg pellet was washed extensively in M9 buffer before extract preparation as described above.

Extracts were size-fractionated on a Superdex 200HR10/30 column (Amersham). Fractions of 250 µl were collected. We used the following size markers: thyroglobulin (669 kDa), ferritin (440 kDa), catalase (232 kDa), aldolase (158 kDa) and albumin (67 kDa).

Expression and purification of full-length Tudor-SN

The complementary DNA was amplified with appropriate restriction sites, cloned into pENTR11 (Invitrogen), and transferred by means of Gateway (Invitrogen) to pDEST17, which has a 6×His N-terminal tag. The plasmid was transformed into BL21 pLysS (Stratagene). Cells were grown to a density of ~0.5 and induced using 1 mM IPTG for 1 h. Cells were freeze-thawed for lysis, suspended in 300 mM NaCl and 50 mM NaPO₄ (pH 7), spun for 20 min at 20,000 g to clear, and loaded onto a HR 10/10 Talon column (Clontech). The peak fractions were diluted with five volumes of buffer B (buffer A plus 2 mM CaCl₂) and eluted from Source Q on a gradient of buffer B from 0 to 1 M KCl. The peak from Q was chromatographed in the same manner on Source S, and finally purified by gel-filtration on Superdex 200, the column being developed using buffer B with 150 mM KCl.

Antibodies

Peptides were synthesized corresponding to eight amino acids from the N and C termini of FXR, VIG and Tudor-SN. Peptides were conjugated to KLH (Pierce) and injected into rabbits. The antisera were affinity-purified using agarose-coupled peptide (Pierce). The FMRP antibody was purchased from Covance.

Immunoprecipitation

For *Drosophila* immunoprecipitations, cells were lysed in 150 mM NaCl, 0.1% NP-40, 20 mM HEPES (pH 7.0), 2 mM MgCl₂, 2 mM CaCl₂, 1 mM DTT and protease inhibitors (Roche). Primary complexes were formed for 1–4 h, and immunoprecipitates were collected using a mix of Protein A–agarose (Pierce) and Protein G–agarose (Roche). Washes were with the same buffer, except that NaCl was adjusted to 400 mM. For HeLa and HEK 293 immunoprecipitation, the buffer was modified to contain 0.5% NP-40, and washes were in the same buffer. For immunoprecipitation–western experiments, cells were previously transfected with siRNAs (Invitrogen). *C. elegans* immunoprecipitations were performed using S100H extract. Extract was incubated with antibody and protein A/G Plus-agarose beads (Santa Cruz) for a period ranging from 5 h to overnight. After extensive washing, the beads were put in SDS–polyacrylamide gel electrophoresis loading buffer, or RNA was isolated using Trizol.

Fractionation of nuclei and cytoplasm

For *Drosophila* cells, fractionation was as described in ref. 28. For 293 cells, fractionation was performed as described in ref. 29. Histone H3 antibody was purchased from Upstate; dynein IC antibody was purchased from Covance.

Immunodepletion

Purified Tudor-SN fractions from Superdex were subjected to two rounds of immunoprecipitation using approximately 10 µg of affinity-purified antibody and protein A/G beads.

Immunofluorescence

Immunofluorescence was carried out using a modification of the protocol described in ref. 30. The concentration of paraformaldehyde used was 0.5%.

Received 23 May; accepted 21 July 2003; doi:10.1038/nature01956.

1. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–296 (2000).

- Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
- Caudy, A. A., Myers, M., Hannon, G. J. & Hammond, S. M. Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev.* **16**, 2491–2496 (2002).
- Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
- Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950–952 (1999).
- Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* **20**, 6877–6888 (2001).
- Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
- Maurer-Stroh, S. *et al.* The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem. Sci.* **28**, 69–74 (2003).
- Ponting, C. P. P100, a transcriptional coactivator, is a human homologue of staphylococcal nuclease. *Protein Sci.* **6**, 459–463 (1997).
- Calabault, I. & Mornon, J. P. The human EBNA-2 coactivator p100: Multidomain organization and relationship to the staphylococcal nuclease fold and to the tudor protein involved in *Drosophila melanogaster* development. *Biochem. J.* **321**, 125–132 (1997).
- Tong, X., Drapkin, R., Yalamanchili, R., Mosialos, G. & Kieff, E. The Epstein–Barr virus nuclear protein 2 acidic domain forms a complex with a novel cellular coactivator that can interact with TFIIE. *Mol. Cell. Biol.* **15**, 4735–4744 (1995).
- Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: Tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* **16**, 2733–2742 (2002).
- Hutvagner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056–2060 (2002).
- Mourelatos, Z. *et al.* miRNPs: A novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16**, 720–728 (2002).
- Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563–574 (2002).
- Keenan, T. W., Winter, S., Rackwitz, H. R. & Heid, H. W. Nuclear coactivator protein p100 is present in endoplasmic reticulum and lipid droplets of milk secreting cells. *Biochim. Biophys. Acta* **1523**, 84–90 (2000).
- Serpensu, E. H., Shortle, D. & Mildvan, A. S. Kinetic and magnetic resonance studies of active-site mutants of staphylococcal nuclease: Factors contributing to catalysis. *Biochemistry* **26**, 1289–1300 (1987).
- Cuatrecasas, P., Fuchs, S. & Anfinsen, C. B. The binding of nucleotides and calcium to the extracellular nuclease of *Staphylococcus aureus*. Studies by gel filtration. *J. Biol. Chem.* **242**, 3063–3067 (1967).
- Slack, F. J. *et al.* The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. *Mol. Cell* **5**, 659–669 (2000).
- Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
- Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
- Knight, S. W. & Bass, B. L. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* **293**, 2269–2271 (2001).
- Reinhart, B. J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
- Llave, C., Xie, Z., Kasschau, K. D. & Carrington, J. C. Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**, 2053–2056 (2002).
- Tang, G., Reinhart, B. J., Bartel, D. P. & Zamore, P. D. A biochemical framework for RNA silencing in plants. *Genes Dev.* **17**, 49–63 (2003).
- Spankuch-Schmitt, B., Bereiter-Hahn, J., Kaufmann, M. & Strebhardt, K. Effect of RNA silencing of polo-like kinase-1 (PLK1) on apoptosis and spindle formation in human cancer cells. *J. Natl Cancer Inst.* **94**, 1863–1877 (2002).
- Capodici, J., Kariko, K. & Weissman, D. Inhibition of HIV-1 infection by small interfering RNA-mediated RNA interference. *J. Immunol.* **169**, 5196–5201 (2002).
- Hirayoshi, K. & Lis, J. T. Nuclear run-on assays: assessing transcription by measuring density of engaged RNA polymerases. *Methods Enzymol.* **304**, 351–362 (1999).
- Avalos, R. T., Yu, Z. & Nayak, D. P. Association of influenza virus NP and M1 proteins with cellular cytoskeletal elements in influenza virus-infected cells. *J. Virol.* **71**, 2947–2958 (1997).
- Spector, D. L. & Smith, H. C. Redistribution of U-sRNPs during mitosis. *Exp. Cell Res.* **163**, 87–94 (1986).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Mildvan, M. Tijsterman and T. Sijen for discussions. We thank T. Keenan for an anti-p100 antibody, T. Hobman for GERP (EIF2C2/hAgo2) antibody, H. Siomi for a FXR monoclonal antibody, and F. Slack for the *lacZ-lin-41* reporter. A.A.C. is a George A. and Marjorie H. Anderson Fellow of the Watson School of Biological Sciences, and a Howard Hughes Medical Institute Predoctoral Fellow. A.M.D. is a David Koch Fellow of the Watson School of Biological Sciences. J.M.S. is supported by a postdoctoral fellowship from the US Army Prostate Cancer Research programme. G.J.H. is a Rita Allen Foundation Scholar and is supported by an Innovator Award from the US Army Breast Cancer Research programme. This work was also supported by a grant from the National Institutes of Health (G.J.H.) and by a VENI fellowship from the Netherlands Organization for Scientific Research (RFK).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.J.H. (hannon@cshl.org) or R.H.A.P. (plasterk@niob.knaw.nl).

The nuclear RNase III Drosha initiates microRNA processing

Yoontae Lee¹, Chiyoung Ahn¹, Jinju Han¹, Hyeonjeong Choi¹, Jaekwang Kim¹, Jeongbin Yim¹, Junho Lee², Patrick Provost³, Olof Rådmark⁴, Sunyoung Kim¹ & V. Narry Kim¹

¹Institute of Molecular Biology and Genetics and School of Biological Sciences, Seoul National University, Seoul 151-742, Korea

²Department of Biology, Yonsei University, Seoul 120-749, Korea

³Centre de Recherche en Rhumatologie et Immunologie, Centre de Recherche du CHUL, Ste-Foy, Quebec, Canada G1V 4G2

⁴Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm S-171 77, Sweden

Hundreds of small RNAs of ~22 nucleotides, collectively named microRNAs (miRNAs), have been discovered recently in animals and plants^{1–10}. Although their functions are being unravelled^{1,2,11–13}, their mechanism of biogenesis remains poorly understood. miRNAs are transcribed as long primary transcripts (pri-miRNAs) whose maturation occurs through sequential processing events: the nuclear processing of the pri-miRNAs into stem-loop precursors of ~70 nucleotides (pre-miRNAs), and the cytoplasmic processing of pre-miRNAs into mature miRNAs¹⁴. Dicer, a member of the RNase III superfamily of bidentate nucleases, mediates the latter step^{15–19}, whereas the processing enzyme for the former step is unknown. Here we identify another RNase III, human Drosha, as the core nuclease that executes the initiation step of miRNA processing in the nucleus. Immunopurified Drosha cleaved pri-miRNA to release pre-miRNA *in vitro*. Furthermore, RNA interference of Drosha resulted in the strong accumulation of pri-miRNA and the reduction of pre-miRNA and mature miRNA *in vivo*. Thus, the two RNase III proteins,

Drosha and Dicer, may collaborate in the stepwise processing of miRNAs, and have key roles in miRNA-mediated gene regulation in processes such as development and differentiation.

We have previously shown that miRNA genes are expressed as primary transcripts (pri-miRNAs) that are longer than pre-miRNAs¹⁴. Pri-miRNAs are trimmed into ~70-nucleotide (nt) pre-miRNA forms, mainly in the nucleus^{14,20}. After this initial processing, the pre-miRNAs are exported to the cytoplasm, and are cleaved to generate the final products of ~22 nt by Dicer^{15–19}. Although Dicer-mediated processing has been studied intensively, the initiation step remains uncharacterized.

To understand the mechanism of the initial processing, we first determined the sites of cleavage by mapping the 5' and 3' ends of pre-miRNA. Pre-miR-30a was cloned by the directional cloning method originally described by Tuschl and others²¹. Briefly, pre-miR-30a was prepared by *in vitro* processing of pri-miRNA labelled with low specific activity and ligated to 5' and 3' adaptor molecules, followed by RT-PCR (polymerase chain reaction with reverse transcription), cloning and sequencing (Supplementary Information). The result showed that pre-miR-30a is a stem-loop of 63 nt (Fig. 1a). Interestingly, the stem-loop contains a 2-nt overhang at its 3' end, which is a characteristic of the products of the RNase-III-mediated cleavage reaction. It is also intriguing that the termini of pre-miR-30a are identical to those of mature miR-30a and miR-30a*, indicating that the initial processing may predetermine the sequences of the final products.

To further characterize the nuclease activity towards pri-miRNA, *cis*-acting requirements for processing were examined by a series of mutations on miR-30a. Deletional mutagenesis followed by *in vitro* processing showed that the sequences covering 20 nt upstream and 25 nt downstream from the cleavage sites are sufficient to support processing (Fig. 1a and Supplementary Information), indicating that the *cis*-acting elements reside in the immediate vicinity and/or inside stem-loop pre-miRNA. Site-directed mutagenesis was then

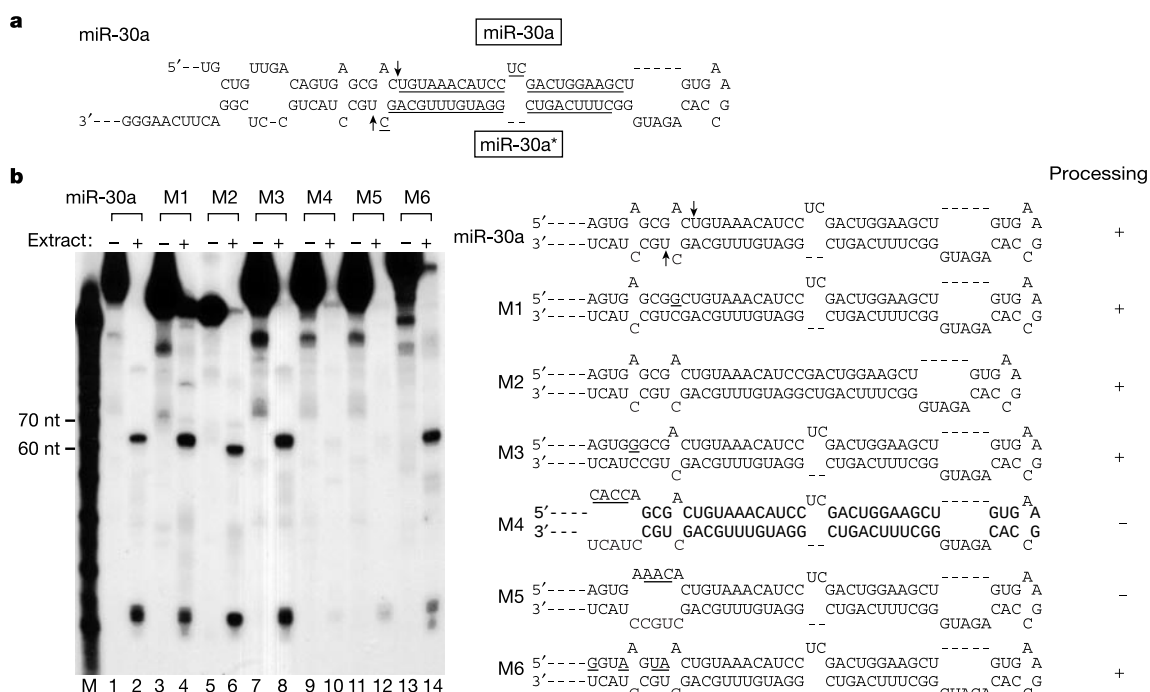


Figure 1 Cleavage sites and *cis*-acting requirements for the initiation step of miRNA processing. **a**, Sequences of the miR-30a precursor that are sufficient to support processing *in vitro*. The cleavage sites (the 5' and 3' ends of pre-miR-30a) are indicated by arrows. The sequences for mature miR-30a and miR-30a* are underlined. **b**, Site-

directed mutagenesis of pre-miR-30a. Altered sequences are underlined. Note that all of the precursors contain the sequences covering 20 nt upstream and 25 nt downstream from the cleavage sites, as in **a**. Size markers (Decade marker, Ambion) were loaded on lane M.

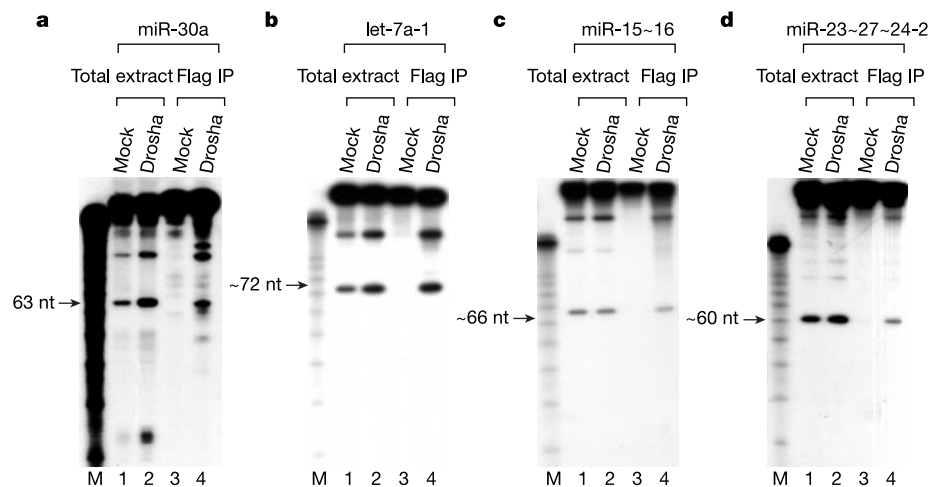


Figure 2 *In vitro* processing of miRNA by Drosha. Total-cell extract (Total extract) or immunoprecipitate (Flag IP) from mock-transfected (Mock) or Drosha–Flag-transfected HEK 293T cells (Drosha) were used for *in vitro* processing. **a**, miR-30a. **b**, let-7a-1. **c**, miR-15~16. **d**, miR-23~27~24-2.

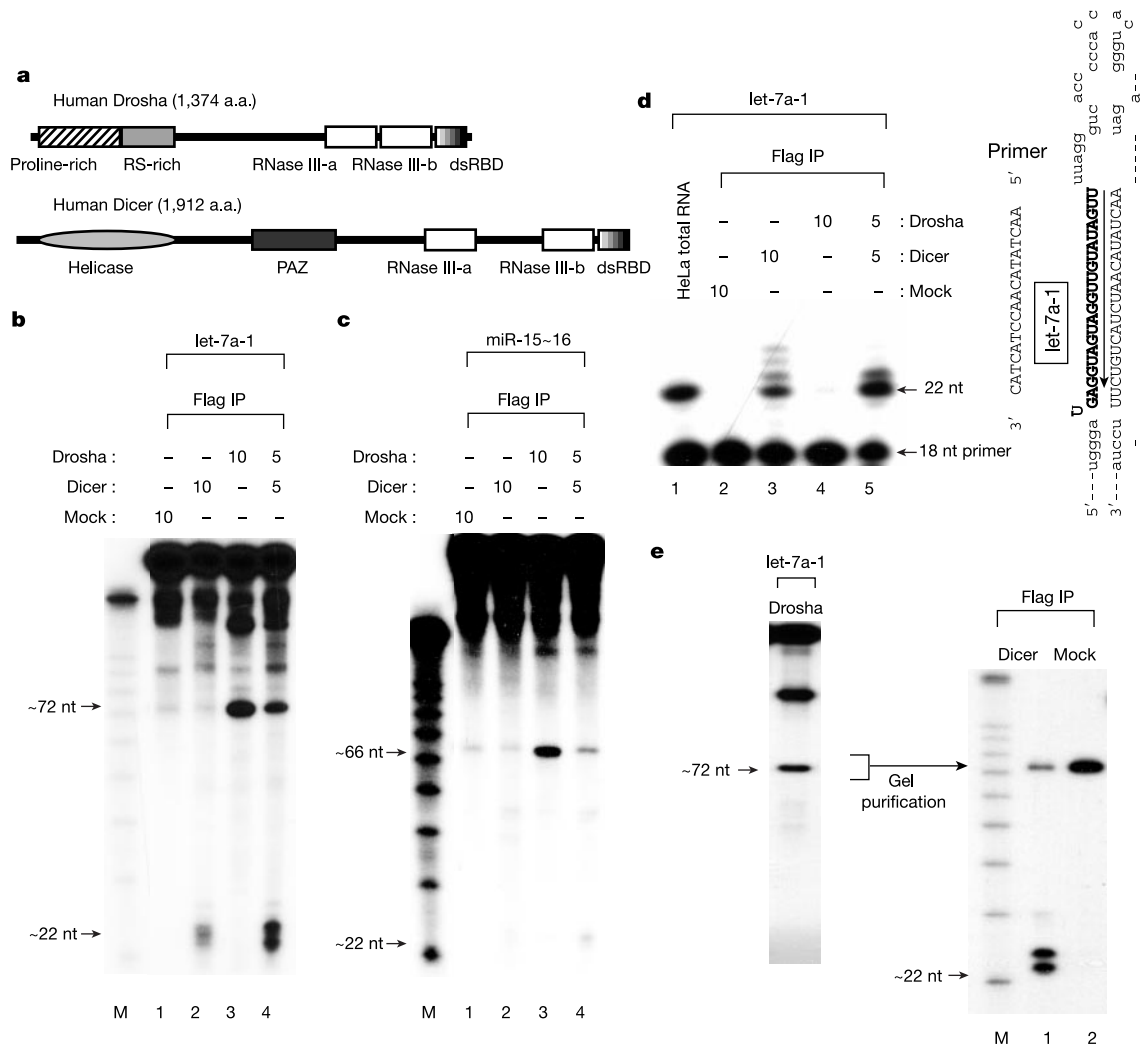


Figure 3 *In vitro* processing activities of Drosha and Dicer. **a**, Domain organization of Drosha and Dicer. **b**, *In vitro* processing of let-7a-1 using immunopurified Drosha and Dicer proteins. Indicated amounts of beads containing immunoprecipitated proteins (μ l) were used for *in vitro* processing. **c**, *In vitro* processing of miR-15~16. **d**, Primer extension analysis using the HeLa total RNAs and the *in vitro* processing product. **e**, Stepwise processing of let-7a-1 by Drosha and Dicer. The indicated fragment from the *in vitro* processing using Drosha was gel-purified and incubated with Dicer.

carried out to remove the internal loops and bulges at or near the cleavage sites (M1, M2 and M3 in Fig. 1b). These mutants were processed efficiently, indicating that the internal loops (M1 and M3) and the bulge (M2) are not essential for processing. On the contrary, mutations that disturbed the stem structure (M4 and M5) markedly reduced the efficiency of the initial processing, whereas a mutant with altered sequences that retained the double-stranded (ds) structure (M6) made a good substrate for processing. This is consistent with a recent study in which the *cis*-acting element of miR-30a was mapped to the first stem outside the cleavage sites²⁰. These results demonstrate the requirement for dsRNA structure around the cleavage sites, supporting the notion that the nuclease of interest may belong to the RNase III family. Members of the RNase III family are known to require divalent cations, including Mg²⁺, for catalysis. It is noted that MgCl₂ was needed for the initial processing of miRNAs *in vitro* (data not shown).

RNase III proteins are dsRNA-specific endonucleases that introduce staggered cuts on each side of the RNA helix²². They are grouped into three classes on the basis of domain organization. Class I enzymes found in bacteria and yeasts contain one conserved RNase III domain and an adjacent dsRNA-binding domain

(dsRBD). Class II enzymes, including Drosha and its homologues, have tandem RNase III domains and one dsRBD, as well as an extended amino-terminal domain of unknown function^{23,24}. Dicer homologues in diverse species are members of class III and contain a putative helicase domain and a PAZ domain, in addition to tandem nuclease domains and a dsRBD. Human database searches revealed three RNase-III-type proteins. L44 contains a single RNase III domain and is a component of the large subunit of the mitochondrial ribosome²⁵. The second RNase III, Dicer, is a cytoplasmic (endoplasmic reticulum–cytosol interface) protein^{26,27} whose function is well established in the late stage of miRNA processing^{15–19}. The third RNase III, Drosha, previously implicated in pre-rRNA (ribosomal RNA) processing, localizes predominantly to the nucleus²⁸, making it a good candidate for the nuclear processing factor for miRNAs.

To examine this possibility, Drosha protein tagged with the Flag epitope was expressed by transfection into HEK 293T cells and isolated by immunoprecipitation from the total-cell lysate using antibody-conjugated agarose. The precipitated protein was then used in an *in vitro* processing reaction. Treatment of the primary precursors with Drosha immunoprecipitate yielded fragments of 60–70 nt in length (Fig. 2), which corresponds to pre-miRNAs¹⁴. Therefore, Drosha can catalyse the initiation step of miRNA processing *in vitro*.

To analyse the processing activities of the two RNase-III-type proteins, Drosha and Dicer proteins were prepared by immunoprecipitation and used for *in vitro* processing. Drosha produced ~70-nt fragments, whereas Dicer cleaved the input RNA to ~22-nt fragments (Fig. 3b, c), showing that the two enzymes have different substrate specificity and mechanisms of action. To confirm the identity of the ~22-nt fragments, primer-extension analysis was carried out (Fig. 3d). Extension products of ~22 nt were detected from the *in vitro* processed products (Fig. 3d, lanes 3 and 5) as well as from HeLa total RNA (Fig. 3d, lane 1), verifying that the ~22-nt fragments included the authentic mature let-7a-1. Intriguingly, the final products appear to be more homogeneous with a combination of Drosha and Dicer (Fig. 3d lane 5) than with Dicer alone (lane 3), indicating that processing mediated by two enzymes may increase its accuracy. It is also interesting that a combination of Drosha and Dicer resulted in a higher yield of the final product of ~22 nt than when twice the amount of Dicer alone was applied (Fig. 3b–d). Thus, Drosha and Dicer may work together in a synergistic manner to increase the production of miRNA. Drosha and Dicer were not co-immunoprecipitated (data not shown), arguing against the possibility that they form a heterodimer to function coordinately. To determine whether Drosha acts upstream of Dicer, the Drosha product was gel-purified and used for a further round of *in vitro* processing using Dicer (Fig. 3e). The ~72-nt Drosha product from pri-let-7a-1 was processed to generate the ~22-nt fragments in the presence of Dicer. Therefore, processing can occur in a stepwise manner *in vitro*. Our data suggest that Drosha may generate optimal substrates for Dicer, and that stepwise processing may have certain advantages over Dicer by only operating in terms of accuracy and efficiency. Stepwise processing is likely to be prevalent *in vivo* owing to the separate localization of the two enzymes: Drosha localizes mainly to the nucleus²⁸, and Dicer is found mainly in the cytoplasmic compartment^{26,27}. Consistent with the previous report²⁸, Drosha was immunoprecipitated mainly from the nuclear fraction, and nuclear Drosha catalysed the initiation step of processing *in vitro* (Supplementary Information). Thus, maturation of miRNA may be carried out consecutively by two RNase III proteins, first by Drosha in the nucleus, and later by Dicer in the cytoplasm (for a model, see Supplementary Information).

To confirm the role of Drosha in miRNA maturation, we carried out RNA interference (RNAi) to deplete Drosha in cultured cells, and examined the effect on endogenous miRNA biogenesis. Out of four different short interfering RNA (siRNA) duplexes designed to

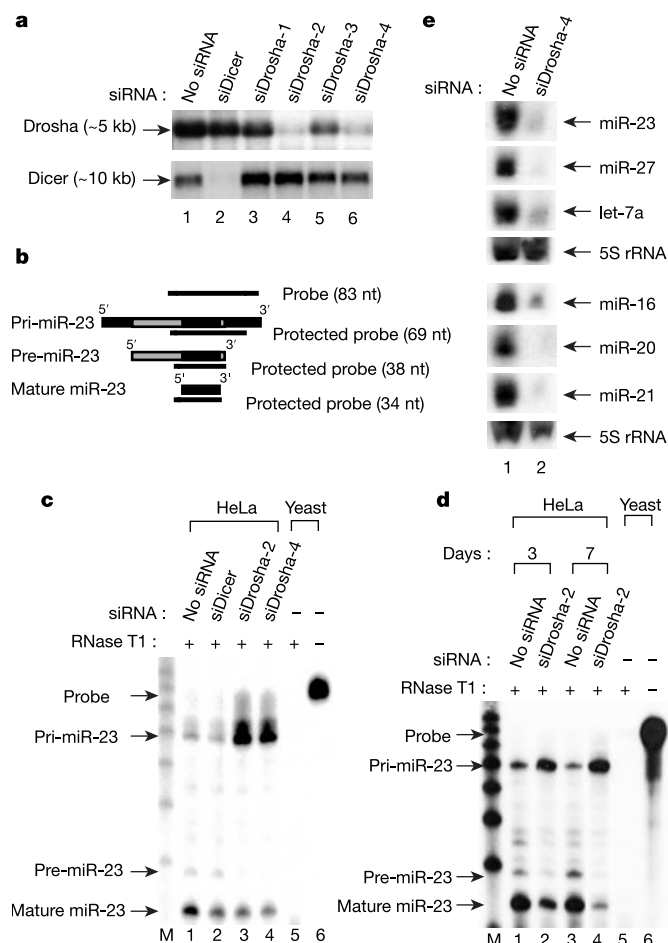


Figure 4 *In vivo* function of Drosha. **a**, Northern blot analysis of Drosha and Dicer messenger RNA, following RNAi in HeLa cells for three days. **b**, RNase protection assay. Three miRNA species, probe and protected fragments are presented schematically. **c**, One-hundred micrograms of total RNA from siRNA transfected HeLa cells were used for RPA. The cells were incubated for three days after transfection. **d**, After incubation for three or seven days after siRNA transfection, total RNA was prepared and 45 µg of total RNA was used for RPA. For the seven-day incubation, the cells were transfected twice. **d**, Northern blot analysis. Total RNA was prepared after double transfection and incubation for seven days. As a loading control, the blots were re-probed for 5S rRNA.

combat Drosha, two duplexes efficiently reduced the level of Drosha mRNA as shown by northern blot analysis (Fig. 4a). A control siRNA duplex targeting Dicer¹⁵ also resulted in effective degradation of Dicer mRNA. Three days after siRNA transfection, an RNase protection assay (RPA) was performed to measure the steady-state level of three miRNA species: pri-, pre- and mature miRNA (Fig. 4b, c). From HeLa total RNA, three specific bands were detected that represented pri-miR-23, pre-miR-23 and mature miR-23. When Drosha was depleted, pri-miR-23 strongly accumulated whereas pre-miR-23 was diminished (Fig. 4c, lanes 3 and 4). We conclude that Drosha is required for the processing of pri-miR-23 into pre-miR-23. The knockdown of Dicer caused a reduction in mature miR-23, as expected (Fig. 4c, lane 2), but pre-miR-23 did not accumulate significantly under our experimental conditions, possibly because pre-miR-23 is unstable.

Mature miR-23 was reduced after Drosha knockdown, as expected (Fig. 4c), and it was further diminished after double transfection of siRNA and prolonged incubation for seven days (Fig. 4d). To show that our observations are not restricted to a certain RNA substrate, we performed northern blot analysis on six randomly chosen miRNAs: miR-23, miR-27, let-7a-1, miR-16, miR-20 and miR-21 (Fig. 4e). miR-30a was not included because of the low abundance of endogenous miR-30a (data not shown), as previously reported²⁹. All the tested miRNAs significantly decreased after Drosha RNAi, suggesting that Drosha may be widely used for the maturation of most, if not all, miRNAs. We observed only a mild growth defect in cells (about 10%) after seven days, indicating that, under our assay conditions, blockade of miRNA processing precedes overall cellular malfunction and death. Drosha has previously been suggested to function in rRNA processing, on the basis of the observation that the depletion of Drosha using antisense oligonucleotides correlated with the accumulation of 12S and 32S pre-rRNAs²⁸. Although the accumulation of pre-rRNA was modest, and the reduction in the final product, 5.8S rRNA, was not evident²⁸, we do not exclude the possibility that Drosha may have dual functions in both rRNA and miRNA processing.

Our study introduces a new player in miRNA biogenesis. As Drosha homologues are found in *C. elegans*, *D. melanogaster*, mice and humans^{23,24}, the stepwise processing of miRNAs mediated by Drosha and Dicer is likely to be conserved, at least in animals. Stepwise processing might be beneficial in terms of efficiency and accuracy of processing (Fig. 3). In addition, stepwise processing and compartmentalization might allow the fine regulation of miRNA biogenesis at multiple steps. It would be of great interest to understand how, and by what, miRNA-processing enzymes are regulated. Our study also reveals interesting differences between class II and III RNase III proteins in their substrate specificities. It is an open question how Drosha and Dicer recognize their targets in a specific manner when the primary sequences of diverse miRNAs show no conserved elements. In view of this, it will be important to determine structures of miRNA precursors and RNase III proteins. Further studies of Drosha will help to elucidate the action mechanism of RNase III, and to unravel the mechanism of miRNA biogenesis, the novel family of developmental regulators. □

Methods

Immunoprecipitation and *in vitro* processing

Proteins were expressed in transfected HEK 293T cells. Forty-eight hours after transfection, total extract was prepared in Buffer D (20 mM HEPES-KOH, pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSF, 5% glycerol) by sonication followed by centrifugation. Total extract was used for immunoprecipitation in Buffer D using anti-Flag M2 affinity gel (Sigma). After washing, the beads were drained and used for *in vitro* processing¹⁴. Thirty microlitres of processing reaction contained 15 µl of whole-cell extract or the beads from immunoprecipitation, 6.4 mM MgCl₂, 1 unit µl⁻¹ of RNase Inhibitor (TAKARA), and the labelled transcripts of 1 × 10⁴–1 × 10⁵ c.p.m. The reaction mixture was incubated at 37 °C for 90 min.

Primer-extension analysis

Cold let-7a-1 precursor was synthesized *in vitro* and incubated with protein

immunoprecipitates at 37 °C for 90 min. After phenol extraction and ethanol precipitation, the RNA was used for hybridization to 5'-end-labelled primer at 37 °C for 40 min, and subsequently for extension using SUPERScript II (GibcoBRL) at 50 °C for 30 min. Forty-five micrograms of HeLa total RNA were used as a control. Primer sequences are 5'-aactatacaactactac-3'.

RNA interference

siRNA duplexes, siDrosha-1, siDrosha-2, siDrosha-3 and siDrosha-4 (Dharmacon), were designed to anneal to Drosha mRNA in which the target sequences were 5'-AACCGAA GAUACCAUCUCUG-3', 5'-AACGAGUAGGCUUCUGACUU-3', 5'-AAUCGGACC UUCGACAGCAA-3' and 5'-AAGGACCAAGUAUUCAGCAAG-3', respectively. A siRNA duplex for Dicer has previously been described¹⁵. siRNA duplexes were transfected into HeLa cells using Oligofectamine Reagent (Invitrogen)³⁰. Total RNA was prepared three days after transfection and used for northern blot analysis or RPA¹⁴. For double transfection, cells were split again three days after the first transfection, transfected once more on the fourth day, and incubated for three more days. The cells were incubated for seven days in total from the time of the first transfection.

Received 4 May; accepted 25 July 2003; doi:10.1038/nature01957.

- Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
- Reinhart, B. J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853–858 (2001).
- Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858–862 (2001).
- Lee, R. C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862–864 (2001).
- Mourelatos, Z. *et al.* miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16**, 720–728 (2002).
- Llave, C., Kasschau, K. D., Rector, M. A. & Carrington, J. C. Endogenous and silencing-associated small RNAs in plants. *Plant Cell* **14**, 1605–1619 (2002).
- Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, D. P. MicroRNAs in plants. *Genes Dev.* **16**, 1616–1626 (2002).
- Lagos-Quintana, M. *et al.* Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**, 735–739 (2002).
- Park, W., Li, J., Song, R., Messing, J. & Chen, X. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12**, 1484–1495 (2002).
- Llave, C., Xie, Z., Kasschau, K. D. & Carrington, J. C. Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**, 2053–2056 (2002).
- Tang, G., Reinhart, B. J., Bartel, D. P. & Zamore, P. D. A biochemical framework for RNA silencing in plants. *Genes Dev.* **17**, 49–63 (2003).
- Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**, 25–36 (2003).
- Lee, Y., Jeon, K., Lee, J. T., Kim, S. & Kim, V. N. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* **21**, 4663–4670 (2002).
- Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293**, 834–838 (2001).
- Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
- Grishok, A. *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**, 23–34 (2001).
- Ketting, R. F. *et al.* Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* **15**, 2654–2659 (2001).
- Knight, S. W. & Bass, B. L. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* **293**, 2269–2271 (2001).
- Zeng, Y. & Cullen, B. R. Sequence requirements for micro RNA processing and function in human cells. *RNA* **9**, 112–123 (2003).
- Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200 (2001).
- Zamore, P. D. Thirty-three years later, a glimpse at the ribonuclease III active site. *Mol. Cell* **8**, 1158–1160 (2001).
- Filippov, V., Solovoyev, V., Filippova, M. & Gill, S. S. A novel type of RNase III family proteins in eukaryotes. *Gene* **245**, 213–221 (2000).
- Fortin, K. R., Nicholson, R. H. & Nicholson, A. W. Mouse ribonuclease III. cDNA structure, expression analysis, and chromosomal location. *BMC Genomics* **3**, 26 (2002).
- Koc, E. C. *et al.* The large subunit of the mammalian mitochondrial ribosome. Analysis of the complement of ribosomal proteins present. *J. Biol. Chem.* **276**, 43958–43969 (2001).
- Billy, E., Brondani, V., Zhang, H., Muller, U. & Filipowicz, W. Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines. *Proc. Natl Acad. Sci. USA* **98**, 14428–14433 (2001).
- Provost, P. *et al.* Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J.* **21**, 5864–5874 (2002).
- Wu, H., Xu, H., Miraglia, L. J. & Crooke, S. T. Human RNase III is a 160-kDa protein involved in preribosomal RNA processing. *J. Biol. Chem.* **275**, 36957–36965 (2000).
- Zeng, Y., Wagner, E. J. & Cullen, B. R. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol. Cell* **9**, 1327–1333 (2002).
- Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to members of our laboratory and to K. Mitrophanous, J. M. Park and H. E. Kim for their critical reading of this manuscript and for discussion. This work was supported by the Korea Research Foundation and the BK21 Research Fellowship from the Ministry of Education and Human Resources Development of Korea.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.N.K. (narrykim@snu.ac.kr).

A structural state of the myosin V motor without bound nucleotide

Pierre-Damien Coureux¹, Amber L. Wells², Julie Ménétrey¹, Christopher M. Yengo², Carl A. Morris², H. Lee Sweeney² & Anne Houdusse¹

¹Structural Motility, Institut Curie CNRS, UMR144, 26 rue d'Ulm, 75248 Paris cedex 05, France

²Department of Physiology, University of Pennsylvania School of Medicine, 3700 Hamilton Walk, Philadelphia, Pennsylvania 19104-6085, USA

The myosin superfamily of molecular motors use ATP hydrolysis and actin-activated product release to produce directed movement and force¹. Although this is generally thought to involve movement of a mechanical lever arm attached to a motor core^{1,2}, the structural details of the rearrangement in myosin that drive the lever arm motion on actin attachment are unknown. Motivated by kinetic evidence that the processive unconventional myosin, myosin V, populates a unique state in the absence of nucleotide and actin, we obtained a 2.0 Å structure of a myosin V fragment. Here we reveal a conformation of myosin without bound nucleotide. The nucleotide-binding site has adopted new conformations of the nucleotide-binding elements that reduce the affinity for the nucleotide. The major cleft in the molecule has closed, and the lever arm has assumed a position consistent with that in an actomyosin rigor complex. These changes have been accomplished by relative movements of the subdomains of the molecule, and reveal elements of the structural communication between the actin-binding interface and nucleotide-binding site of myosin that underlie the mechanism of chemo-mechanical transduction.

Myosin V is a myosin motor that has a number of structural and kinetic features that allow it to act as a two-headed processive motor protein³. The most notable feature is its long tandem repeat of six calmodulin/light-chain-binding sites, which form a long 'lever arm' that allows myosin V to take multiple 36-nm steps along an actin filament without detachment (that is, processive movement)^{3,4}. This allows the molecule to function as a vesicle transporter⁵. To achieve processive movement, the rates of key kinetic steps of myosin V are very different from myosin II, which results in each head spending most of its ATPase cycle strongly bound to actin⁶.

The current view of how myosin couples ATP hydrolysis and actin binding to movement is known as the lever arm hypothesis^{1,2}. In essence the proposed mechanism is that nucleotide binding, hydrolysis and product release are all coupled to small movements within the myosin motor core. These movements are amplified and transmitted by a region that has been termed the 'converter'^{7,8} to a lever arm consisting of a light-chain-binding helix and associated light chains. The lever arm further amplifies the motions of the converter into large directed movements⁷⁻⁹. In the absence of actin, ATP hydrolysis occurs, but product release is slow, thus trapping the

lever arm in a primed or pre-power-stroke position. Binding to actin causes release of products, movement of the lever arm, and force generation concomitant with formation of strong binding between myosin and actin (see Supplementary Fig. 1).

Although there is growing evidence for this general scheme¹, the proposed structural details of the motor domain changes are based entirely on high-resolution structures of myosin II in states that bind weakly to actin, and thus do not represent force-generating states². Crystals with either vanadate or AlF₄ plus MgADP reveal a state (the 'transition' state) that probably mimics the pre-power-stroke state of myosin compatible with ATP hydrolysis, and before actin binding^{7,10,11}. The 'near-rigor' state, which has been seen with MgADP, MgATP, ATP analogues or no nucleotide in the active site^{12,13}, has been proposed to reveal the position of the myosin lever arm at the end of the power stroke on release of MgADP (an actomyosin state known as rigor). However, it cannot truly represent such a state as there is considerable kinetic evidence that demonstrates that the near-rigor conformation of myosin II cannot bind strongly to actin without significant structural rearrangements^{2,14}. Thus current crystal structures of myosin II offer no explanation of how actin-induced product release increases the affinity of myosin for actin, and how ATP dissociates the actin-myosin nucleotide-free (rigor) complex after release of phosphate and MgADP.

As previously noted⁶ (see also Supplementary Fig. 2), in the absence of nucleotide an expressed truncated form (truncated after the first calmodulin-binding domain) of myosin V with a bound essential light chain (LC1-sa with the extended amino terminus removed¹⁵) binds rapidly to actin in a concentration-dependent manner, is not temperature dependent, and does not saturate over the actin concentrations examined. The rate of skeletal myosin II actin binding is highly temperature dependent and saturates at a moderate rate and actin concentration¹⁶. This indicates that in the absence of nucleotide and actin, myosin II—which mostly populates the near-rigor conformational state—is in a very different conformation from that in the rigor complex that forms after adding actin. On the other hand, the kinetics of myosin V binding appear to be diffusion limited⁶, implying that myosin V in the absence of nucleotide and actin is in a state that is nearly equivalent to the rigor state formed on the addition of actin.

Crystals of this nucleotide-free myosin V construct diffracted to 2.0 Å. The refined structure is unlike that of any of the myosin structures to date. However, as compared to the near-rigor or pre-power-stroke structures, the lever arm is in a position that is similar to that of the near-rigor state (see Supplementary Fig. 3).

Figure 1a demonstrates that there have been movements of the previously defined⁸ subdomains of the myosin motor that allow a different actin interface, closure of a major cleft (50-kDa cleft) in the molecule, and weakening of nucleotide binding. These subdomains of the motor, which have been proposed to move as units connected by flexible connectors or 'joints', are the N-terminal, upper 50-kDa and lower 50-kDa subdomains, and the converter (to which the lever arm is attached). This structure reveals new conformations of the previously defined connectors⁸ (switch II, relay and SH1 helix) as well as revealing the importance of a fourth connector (previously termed the 'strut'¹⁷), which links the lower and upper 50-kDa subdomains near the actin interface. Precise interactions mediated by these connectors, that are different for each myosin state, allow stabilization of the unique subdomain positions in each state. In our new state, for the first time in any myosin structure, there is a significant movement (25° rotation) of the upper 50-kDa subdomain relative to the N-terminal subdomain. As discussed below, this movement is critical for both rearrangement of the nucleotide-binding pocket and closure of the internal cleft between the two 50-kDa subdomains.

On the basis of attempts to dock the initial and subsequent high-resolution structures of myosin II into lower-resolution cryo-

electron microscopy reconstructions of the actin–myosin complex, it has been predicted that the 50-kDa cleft in the myosin molecule would have to close in order to create the rigor interaction^{18,19}. In our new structure, the actin interface is quite different from those of the near-rigor or transition states, owing to a relative rotation and translation of the upper and lower 50-kDa subdomains resulting in closure of the cleft (Fig. 1a) both near the actin- and nucleotide-binding sites. The closure near the actin-binding site requires a conformational change in a connector between the upper and lower 50-kDa subdomains (the strut) that allows it to interact specifically with both subdomains (Fig. 1b). Changes in the strut also allow a large number of direct interactions between the upper and lower 50-kDa subdomains, including a number of van de Waals interactions and specific side-chain interactions. Mutagenesis studies have demonstrated that alterations in the length of the strut prevent strong binding to actin¹⁷.

Previous evidence for some degree of cleft closure on formation of the actin–myosin rigor complex includes a fluorescence change of an engineered tryptophan (F425W) in the upper 50-kDa subdo-

main of smooth muscle myosin II²⁰. In our structure the corresponding residue, N398, forms hydrogen bonds with residues of both the lower 50-kDa subdomain and the strut (Fig. 1b). It is also known that formation of the actin–myosin rigor complex requires a conformational change in myosin II that results in a net exclusion of water molecules from the motor¹⁴. The cleft closure seen in our structure results in exclusion of a considerable amount of water from the molecule, obliterating the aqueous tunnel seen in the near-rigor structure (Fig. 1a).

Another major prediction based on the first myosin II structure was that there would be an opening of the nucleotide pocket triggered by cleft closure (actin binding) to release MgADP¹². However, the elements that coordinate the nucleotide (P loop, switch I) have not been observed to move significantly, even in myosin structures to date that do not have bound nucleotide¹³. In our myosin V structure, the possibility of high-affinity nucleotide binding has been eliminated in a series of unexpected structural rearrangements, rather than a simple opening of the pocket. The P loop and switch I have moved approximately 6.5 Å apart, destroying

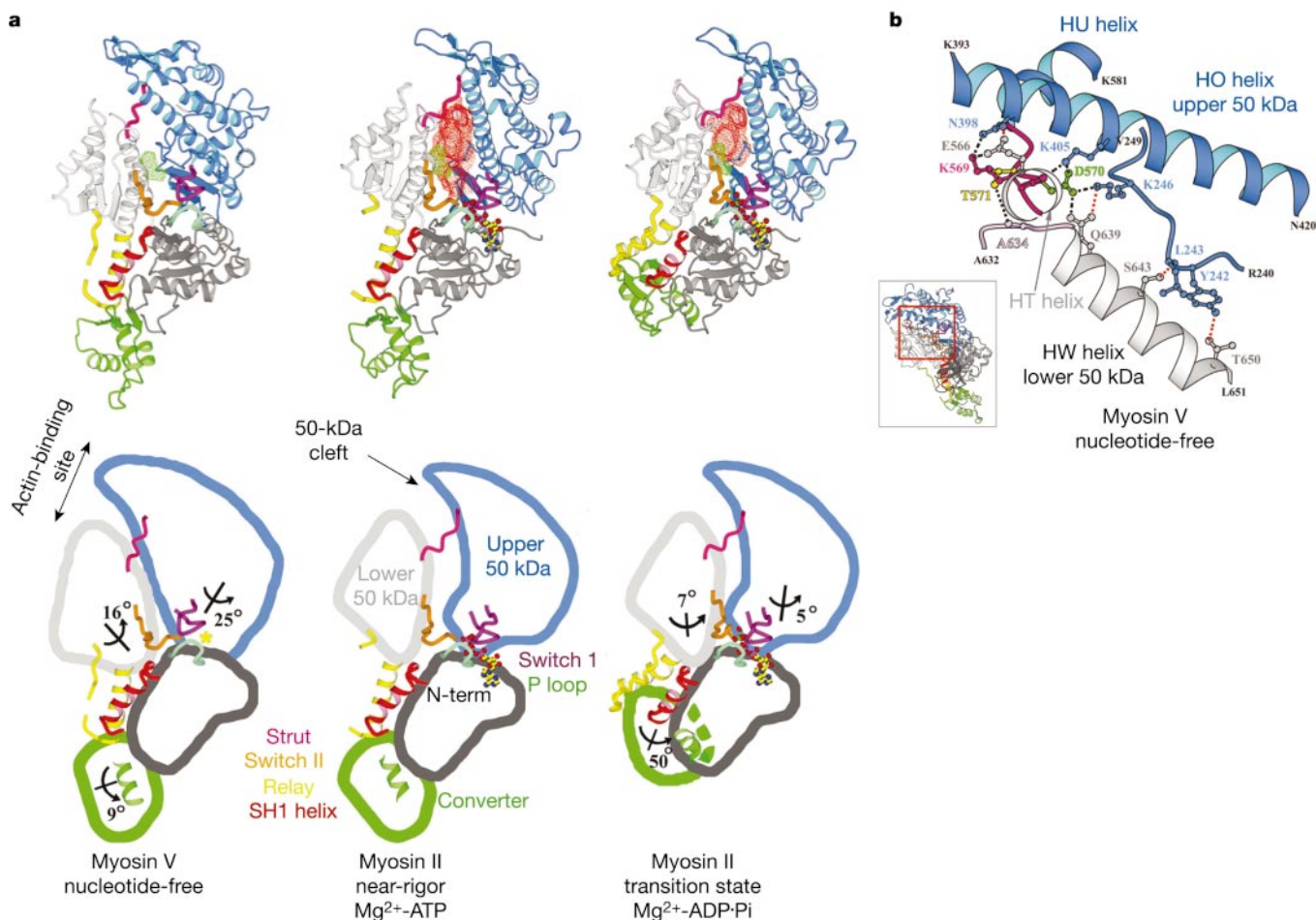


Figure 1 Positions of subdomains and connectors in the three myosin states and closure of the 50-kDa cleft. **a**, A comparison of the myosin V motor domain to the *Dictyostelium* myosin II in the near-rigor and transition states shows the different positions of the subdomains, nucleotide-binding elements and connectors in each state. The structures have been superimposed on the N-terminal subdomains. Relative to this subdomain, the rotation necessary to move from the myosin V state to the near-rigor state is indicated for each subdomain of myosin V; similarly, the rotation necessary to move from the near-rigor to the transition state is indicated on the subdomains of the transition-state structure. Contours of the solvent-accessible cavities for the near-rigor ($1,735 \pm 173 \text{ \AA}^3$) and

transition states ($795 \pm 140 \text{ \AA}^3$) are shown with a red contour, whereas the green contour ($73 \pm 27 \text{ \AA}^3$) depicted in all structures represents an internal cavity of myosin V. Pi, inorganic phosphate. **b**, Shown are specific hydrogen bonds involving the strut that result in cleft closure near the actin interface. Of note is the interaction of D570 with K405 and K246, which are residues conserved in all myosins. However, a number of residues involved in cleft closure in our structure (such as T571, K569, N398) are not absolutely conserved in the myosin superfamily. Variability in cleft interactions could alter the kinetics of cleft closure, and thus the rate of the weak-to-strong binding to actin.

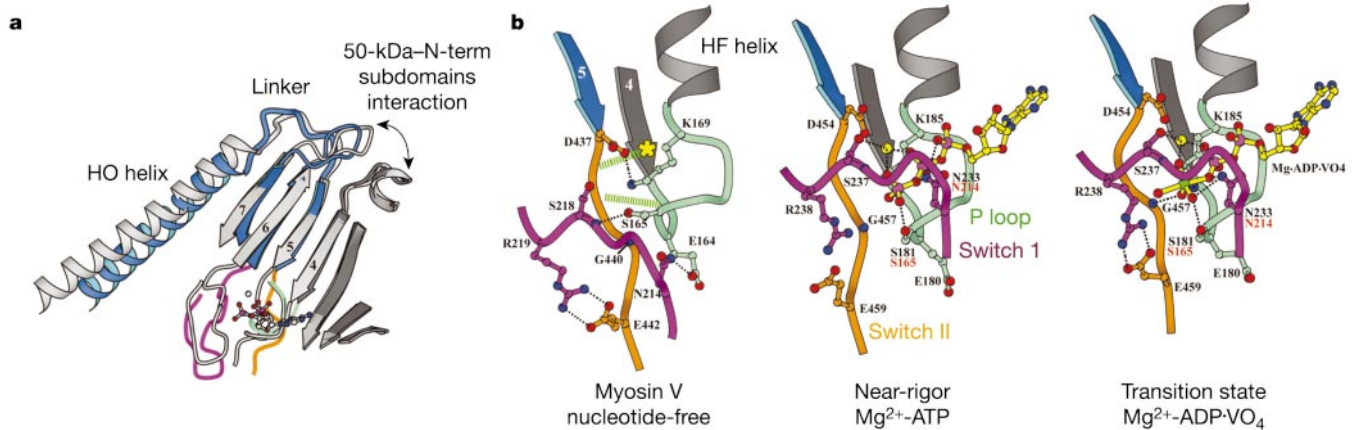


Figure 2 Nucleotide-binding site and distortion of the β -sheet at the interface of the N-terminal and upper 50-kDa subdomains. **a**, Shown is an overlay of the β -sheet (N-terminal subdomain superimposed) between myosin V in blue and near-rigor (*Dictyostelium* myosin II) in grey. Note that strands 5–7, which belong to the upper 50-kDa subdomain are distorted to allow the upper 50-kDa rotation that removes switch I from the nucleotide-binding site. **b**, The positions of the nucleotide-binding elements are shown for

three myosin states. The yellow asterisk in the myosin V structure marks the position of the Mg^{2+} in the other structures. In the transition state, switch II contributes to coordination of the γ -phosphate of the nucleotide, but it bends in myosin V in the opposite direction and forms direct interactions (broken green lines) with the fourth β -strand and the P loop of the N-terminal subdomain.

the ability to coordinate Mg^{2+} and the possibility of switch I interactions with nucleotide. A new conformation of switch II is stabilized by a number of new interactions (Fig. 2b). Note that the switch I conformation is not altered, but simply follows the movement of the upper 50-kDa subdomain relative to the P loop and N-terminal subdomain (Fig. 2a). The position of the P loop itself appears to provide steric hindrance to nucleotide entry (Fig. 2a). However, as its position in our structure is stabilized by weak

interactions, we would anticipate that in solution the P loop would rapidly explore other conformations and would not significantly hinder ATP binding.

A distortion of the seven-stranded β -sheet that couples the N-terminal and upper 50-kDa subdomains is essential to allow the large movement of the upper 50-kDa subdomain (including switch I). Although the amplitude of this distortion is large (Fig. 2a), the key interactions that maintain the sheet have not been altered

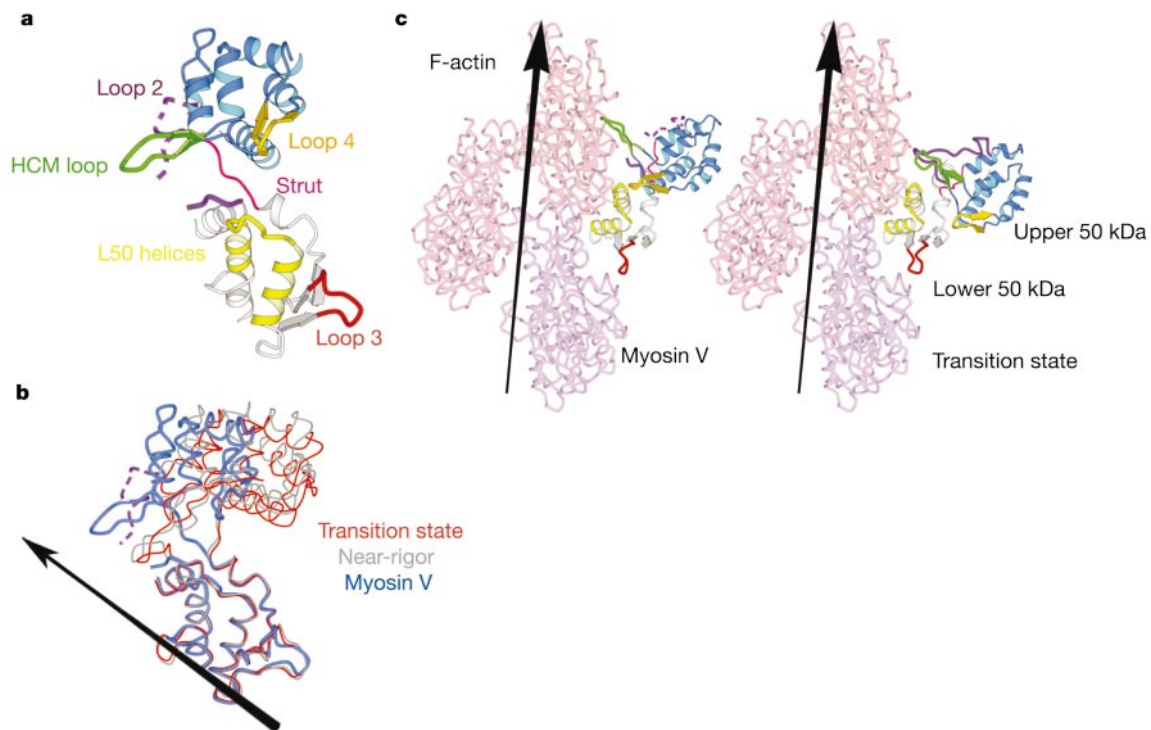


Figure 3 The actin–myosin interface. **a**, Myosin V as viewed from the actin side of the interface reveals the positioning of putative actin-binding elements. **b**, The same view of myosin V is overlaid on the lower 50-kDa subdomains of *Dictyostelium* myosin II structures. Note the conformational change in the strut and the obvious rotation of the upper 50-kDa subdomain towards the actin filament (represented by an arrow) in the

myosin V structure. **c**, The myosin V and transition state actin-binding elements are docked on an actin filament, maintaining the identical positioning of the lower 50-kDa subdomain in both cases. In myosin V, the HCM loop has been repositioned in such a way that it can directly contribute to actin binding.

(unlike what is observed for ARF proteins²¹). The distortion of strands 5–7 of the β -sheet allows switch I and II to follow the rotation of the upper 50-kDa subdomain, while maintaining the upper 50-kDa–N-terminal subdomain interactions on the opposite side of the sheet (Fig. 2a). (A similar β -sheet distortion has also been observed in a new structural state of *Dictyostelium* myosin II (F. J. Kull and D. J. Manstein, personal communication).) This distortion (that is, structural change that does not simply follow the subdomain movement) is greatest for strand 5. That strand connects on one side to switch II, and on the opposite side to a linker coming from the longest helix (HO) of the upper 50-kDa subdomain that begins at the actin interface (HCM loop²², see below). Through this helix, the linker can provide coupling between the β -sheet and the actin interface, and itself undergoes considerable distortion in our new state as compared with either of the weak binding states (transition or near-rigor states).

Switch II has an important role in the positioning of the subdomains that are critical for cleft closure and lever arm position in the new state. Notably, switch II promotes a different set of interactions between the subdomains in the transition state structure. This results in the partial cleft closure of the transition state, as compared to the near-rigor state, which traps the phosphate after ATP hydrolysis and results in repositioning of the lever arm in its pre-power-stroke conformation. However, the interactions that close the cleft in the transition state are not all maintained in our new state. Indeed, part of the region that is closed in the transition state opens to form a small internal cavity in the nucleotide-free structure (Fig. 1a).

If one assumes that the interface between the lower 50-kDa subdomain and actin is essentially the same in all myosin states, then our new structure would form a more extensive interface with actin because the upper 50-kDa subdomain has been moved close to the actin surface. In particular, the HCM loop—which is known to contribute hydrophobic interactions to create strong actin binding²²—and a previously uninvestigated loop (loop 4)¹¹ are now in position to make actin interactions (Fig. 3). Thus our new structure would be predicted to provide a much better fit to either a myosin II or myosin V rigor cryo-electron microscopy map near the actin interface than do any of the existing myosin II near-rigor structures.

On the basis of all of the structural and kinetic evidence described above, we propose that our myosin V structure is a rigor-like state. The positions of the switch elements and the β -sheet seen in our structure explain how cleft closure induced by actin binding results in the weakening of nucleotide affinity. If this rigor-like structure then rearranges to the conformation seen in the near-rigor state on introduction of ATP, it would explain how ATP binding to the actomyosin rigor complex weakens the affinity of myosin for actin. The sequence of events would be initial binding of the phosphates of ATP to the P loop, followed by an inward movement of switch I to coordinate the γ - and β -phosphates and the Mg^{2+} ion. This switch I movement would be mediated by rearrangement of the β -sheet, and accompanied by movement of switch II, and a reopening of the cleft. (The necessity of switch II movement in the transition from rigor to near-rigor is supported by a mutation of switch II (G440A) of myosin V that greatly slows ATP-induced dissociation from actin, while allowing ATP binding²³.) This would decrease the affinity of the myosin for actin, leading to dissociation of the myosin coincident with formation of the near-rigor state. This isomerization and dissociation occurs with minimal movement of the converter and lever arm. Thus ATP binding does not reverse the lever arm movement. Reversal or repriming occurs on isomerization from the near-rigor state to the transition state (while myosin is detached from actin), which immediately precedes ATP hydrolysis.

What cannot be deduced from the myosin V rigor structure are details of the myosin states between the initial weakly bound actin–myosin–ADP·Pi state and the actin–myosin rigor complex. In these

states, phosphate followed by MgADP is released from the myosin. The key interactions between actin and myosin that trigger the conformational changes that lead to release of phosphate and the formation of a strong actin–myosin interface undoubtedly involve a flexible loop that connects the upper and lower 50-kDa subdomains, and is commonly referred to as loop 2 (refs 24, 25). This loop does not contribute any interactions to our new structure, and thus is disordered. Binding to actin probably orders this loop and in doing so destabilizes the myosin·ADP·Pi structure and initiates cleft closure at the actin interface. Indeed, mutations in this loop can greatly slow and even prevent the release of phosphate and formation of strong actin binding in myosin II²⁵.

After phosphate release, myosin V populates a strong actin-binding state, with MgADP bound strongly to the nucleotide-binding site⁶. (This is the predominate steady-state intermediate of myosin V's actin-activated ATPase cycle⁶.) Clearly the alterations in the positions of the switch elements, P loop and β -sheet elements described above cannot occur and maintain tight binding of MgADP, as occurs for myosin V⁶ and other myosin isoforms⁹. Thus another state must exist in which there are changes in the relative positions of the upper and lower 50-kDa subdomains to achieve a high-affinity actin interface without significantly perturbing nucleotide binding.

We assert that myosin V populates a rigor-like conformation in the absence of both actin and nucleotide. This structure, in revealing the nature of the structural communication between the actin- and nucleotide-binding sites, provides an explanation of how strong binding of myosin to actin weakens the affinity of myosin for nucleotide, and vice versa. The major predictions of changes in the myosin structure that must occur for formation of the rigor complex with actin have been realized in this myosin V structure. It is important to note that myosin V is not an anomaly among myosins, as we also have a near-rigor state structure of this same myosin V construct in the presence of MgADP·BeFx (data not shown). Thus it is likely that the conformation we report here is essentially the same for all myosins when bound to actin in the absence of nucleotide, with the exception of the flexible loops at the actin interface that probably become ordered when myosin binds to actin. We refer to this new structural state of myosin as the closed-cleft state. □

Methods

Protein engineering and preparation

Myosin V was expressed using the baculovirus/SF9 cell expression system. The expression and purification were as previously described⁶. To create the recombinant virus used for expression, the complementary DNA coding for chicken myosin V was truncated after the codon corresponding to Arg 792. This construct encompassed the motor domain and the first light-chain/calmodulin-binding site of myosin V. A truncated cDNA for the LC1-sa light chain¹⁵ was co-expressed with the truncated myosin V heavy chain.

Crystallization, X-ray data collection and processing

Crystals were grown in hanging drops by vapour diffusion using equal amounts of reservoir solution (containing 6% PEG 8000, 50 mM MOPS pH 6.5, 2 mM DTT and 2 mM $NaNO_3$) and stock solution of the protein at 8 mg ml^{−1}. The protein/precipitant drops were microseeded the next day by streak seeding from previous crystallizations. The crystals belonged to the $P2_1$ space group (cell dimension: $a = 53.9$ Å, $b = 98.2$ Å, $c = 111.4$ Å and $\beta = 101.4^\circ$, one molecule per asymmetric unit). An X-ray data set was collected up to 2 Å at 100 K at the European Synchrotron Radiation Facility (ESRF) beamline ID-29 ($R_{\text{merge}} = 6.0\%$ (last shell = 33.5%)) with overall completeness 96.8% (2.12–2.05 Å, 88.4%). We processed data using the programs DENZO²⁶ and SCALEPACK²⁶.

Structure refinement

The structure was solved by molecular replacement using the program AmoRe²⁷. The initial model was the motor domain of skeletal striated muscle (Protein Data Bank code 2MYS). Several steps of rigid body fitting performed with AmoRe²⁷ (each subdomain has been considered as a rigid group) were necessary to obtain a solution. We carried out model building and refinement using the program O²⁸ and Refmac5 (ref. 27). The final structure was refined to 2.0 Å (R_{free} value, 26.3%; R_{cryst} value, 21.9%) and validated using the program PROCHECK²⁹. The main-chain dihedral angles for 92% of the non-glycine residues are in the maximum allowed region and none are in the disallowed region of the Ramachandran map. The model has root-mean-square deviations in bond length and bond angles of 0.014 Å and 1.296°, respectively. Diagrams were computed using

MOLSCRIPT³⁰ and with the *Dictyostelium* myosin II structures (Protein Data Bank codes 1FMW and 1VOM). For Fig. 3, docking of myosin V was performed using a cryo-electron microscopy map of myosin II (K. C. Holmes and R. R. Schröder, personal communication and data not shown).

Received 13 March; accepted 18 July 2003; doi:10.1038/nature01927.

- Holmes, K. C. & Geeves, M. A. Structural mechanism of muscle contraction. *Annu. Rev. Biochem.* **68**, 687–728 (1999).
- Houdusse, A. & Sweeney, H. L. Myosin motors: missing structures and hidden springs. *Curr. Opin. Struct. Biol.* **11**, 182–194 (2001).
- Mehta, A. D. E. *et al.* Myosin V is a processive actin-based motor. *Nature* **400**, 590–593 (1999).
- Purcell, T. J., Morris, C., Spudich, J. A. & Sweeney, H. L. Role of the lever arm in the processive stepping of myosin V. *Proc. Natl Acad. Sci. USA* **99**, 14159–14164 (2002).
- Titus, M. A. Myosin V—the multi-purpose transport motor. *Curr. Biol.* **7**, R301–R304 (1997).
- De La Cruz, E. M., Wells, A. L., Safer, D., Ostap, E. M. & Sweeney, H. L. The kinetic mechanism of myosin V. *Proc. Natl Acad. Sci. USA* **96**, 13726–13731 (1999).
- Dominguez, R., Freyzon, Y., Trybus, K. M. & Cohen, C. Crystal structure of a vertebrate smooth muscle myosin motor domain and its complex with the essential light chain: visualization of the pre-power stroke state. *Cell* **94**, 559–571 (1998).
- Houdusse, A., Kalabokis, V. N., Himmel, D., Szent-Gyorgyi, A. G. & Cohen, C. Atomic structure of scallop myosin subfragment S1 complexed with MgADP: a novel conformation of the myosin head. *Cell* **97**, 459–470 (1999).
- Whittaker, M. *et al.* A 35 Å movement of smooth muscle myosin on ADP release. *Nature* **378**, 748–751 (1995).
- Fisher, A. J. *et al.* X-ray structures of the myosin motor domain of *Dictyostelium discoideum* complexed with MgADP·BeF₃⁻ and MgADP·AlF₄⁻. *Biochemistry* **34**, 8960–8972 (1995).
- Kollmar, M., Durrwang, U., Kliche, W., Manstein, D. J. & Kull, F. J. Crystal structure of the motor domain of a class-I myosin. *EMBO J.* **21**, 2517–2525 (2002).
- Rayment, I. *et al.* Three-dimensional structure of myosin subfragment-1: A molecular motor. *Science* **261**, 50–58 (1993).
- Bauer, C. B., Holden, H. M., Thoden, J. B., Smith, R. & Rayment, I. X-ray structures of the apo and MgATP-bound states of *Dictyostelium discoideum* myosin motor domain. *J. Biol. Chem.* **275**, 38494–38499 (2000).
- Coates, J. H., Criddle, A. H. & Geeves, M. A. Pressure-relaxation studies of pyrene-labelled actin and myosin subfragment-1 from rabbit skeletal muscle. *Biochem. J.* **232**, 351–356 (1985).
- De La Cruz, E. M., Wells, A. L., Sweeney, H. L. & Ostap, E. M. Actin and light chain isoform dependence of myosin V kinetics. *Biochemistry* **39**, 14196–14202 (2000).
- Taylor, E. W. Kinetic studies on the association and dissociation of myosin subfragment 1 and actin. *J. Biol. Chem.* **266**, 294–302 (1991).
- Sasaki, N., Ohkura, R. & Sutoh, K. Insertion or deletion of a single residue in the strut sequence of *Dictyostelium* myosin II abolishes strong binding to actin. *J. Biol. Chem.* **275**, 38705–38709 (2000).
- Rayment, I. *et al.* Structure of the actin-myosin complex and its implications for muscle contraction. *Science* **261**, 58–61 (1993).
- Volkman, N. *et al.* Evidence for cleft closure in actomyosin upon ADP release. *Nature Struct. Biol.* **7**, 1147–1155 (2000).
- Yengo, C. M., De La Cruz, E. M., Chrin, L. R., Gaffney, D. P. II & Berger, C. L. Actin-induced closure of the actin-binding cleft of smooth muscle myosin. *J. Biol. Chem.* **277**, 24114–24119 (2002).
- Pasqualato, S., Renault, L. & Cherfils, J. Arp, Arl, Arp and Sar proteins: a family of GTP-binding proteins with a structural device for 'front-back' communication. *EMBO Rep.* **3**, 1035–1041 (2002).
- Sasaki, N., Asukagawa, H., Yasuda, R., Hiratsuka, T. & Sutoh, K. Deletion of the myopathy loop of *Dictyostelium* myosin II and its impact on motor functions. *J. Biol. Chem.* **274**, 37840–37844 (1999).
- Yengo, C. M., De la Cruz, E. M., Safer, D., Ostap, E. M. & Sweeney, H. L. Kinetic characterization of the weak binding states of myosin V. *Biochemistry* **41**, 8508–8517 (2002).
- Spudich, J. A. How molecular motors work. *Nature* **372**, 515–518 (1994).
- Joel, P. B., Trybus, K. M. & Sweeney, H. L. Two conserved lysines at the 50/20-kDa junction of myosin are necessary for triggering actin activation. *J. Biol. Chem.* **276**, 2998–3003 (2001).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–325 (1997).
- Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemistry of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
- Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by grants from the National Institutes of Health (H.L.S.), the CNRS and the ARC (A.H.). We thank D. Picot and the staff of the European Synchrotron Radiation Facility for assistance during data collection. We are also grateful to C. Baldacchino and J. Kibbe for help in protein purification.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.H. (anne.houdusse@curie.fr) or H.L.S. (lsweeney@mail.med.upenn.edu). Atomic coordinates have been deposited in the Protein Data Bank under the accession code 1OE9.

Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide

Kenneth C. Holmes, Isabel Angert, F. Jon Kull*, Werner Jahn & Rasmus R. Schröder

Department of Biophysics, Max Planck Institute for Medical Research, 69120 Heidelberg, Germany

* Present address: Dartmouth College, Hanover, New Hampshire 03755, USA

Muscle contraction involves the cyclic interaction of the myosin cross-bridges with the actin filament, which is coupled to steps in the hydrolysis of ATP¹. While bound to actin each cross-bridge undergoes a conformational change, often referred to as the “power stroke”², which moves the actin filament past the myosin filaments; this is associated with the release of the products of ATP hydrolysis and a stronger binding of myosin to actin. The association of a new ATP molecule weakens the binding again, and the attached cross-bridge rapidly dissociates from actin. The nucleotide is then hydrolysed, the conformational change reverses, and the myosin cross-bridge reattaches to actin. X-ray crystallography has determined the structural basis of the power stroke, but it is still not clear why the binding of actin weakens that of the nucleotide and vice versa. Here we describe, by fitting atomic models of actin and the myosin cross-bridge into high-resolution electron cryo-microscopy three-dimensional reconstructions, the molecular basis of this linkage. The closing of the actin-binding cleft when actin binds is structurally coupled to the opening of the nucleotide-binding pocket.

A complex of actin and myosin cross-bridges, formed by mixing fibres of actin with myosin cross-bridges in the absence of ATP (decorated actin), is a model system for the strong binding or ‘rigor’ state. Decorated actin has been studied by electron cryo-microscopy^{3–5}. To obtain higher resolution structural data, we have studied the actin–myosin cross-bridge complex by electron cryo-microscopy with ‘energy-filtering’. We have also used a new system for three-dimensional (3D) reconstruction to extract the maximum possible resolution from the data. Zero-loss energy-filtered imaging markedly improves the signal-to-noise ratio in the micrographs; in turn, this allows an accurate correction for the contrast transfer function and better alignment procedures during image processing, and thus a much improved resolution (14 Å) and a more detailed description of the actin–myosin interaction.

The structure of subfragment 1 (S1) of chicken skeletal myosin without bound nucleotide has been solved by X-ray crystallography⁶. It is a P-loop protein with switch 1 and switch 2 elements that are similar to those of the G proteins. The ATP is bound by the P-loop between the switch 1 and switch 2 elements. In the pre-power stroke state^{7,8}, both switch 1 and switch 2 make specific hydrogen-bond-mediated contact with the polyphosphate moiety of the ATP, particularly with the γ-phosphate. In the rigor state, the switch 2 element swings away.

Here, the actin filament structure was derived by fitting the crystal structure of monomeric actin (G-actin) to X-ray fibre diffraction patterns obtained from orientated gels of actin fibres (F-actin; see Supplementary Information).

The improvement in signal-to-noise ratio arising from using the energy filter allowed us to take data close to focus and at a relatively low electron dose (Fig. 1). By merging focal series at four different degrees of underfocus after correcting for the contrast transfer function⁹, we obtained the density shown as a surface representation in Fig. 2. The resolution of the map was 14 Å (for the resolution

criteria and the reconstruction method, see Methods and Supplementary Information).

The atomic models of the actin filament and of the myosin cross-bridge were fitted into the new electron density (Fig. 2a) by an iterative least-squares procedure. The actin was allowed two degrees of freedom (height and azimuth), the cross-bridge was allowed six. The fit of the actin was excellent (see Supplementary Movies). In addition, no major adjustments to the geometry of the crystal structure of the myosin cross-bridge seemed to be necessary to achieve a moderately good fit. Maps of $\rho_{\text{obs}} - \rho_{\text{calc}}$ showed, however, that the occupancy of the regulatory light chain (RLC) was only 50%. In the subsequent refinement, the occupancy of the RLC was set at 50%.

More detailed inspection showed that the upper domain of relative molecular mass 50,000 (M_r 50K) protruded out of the density (Fig. 2a, b). Pairs of positive and negative peaks in the $\rho_{\text{obs}} - \rho_{\text{calc}}$ map also showed that a shift of the upper 50K domain would be necessary to fit the density (see Supplementary Movies). In a subsequent least-squares fit, we therefore allowed the upper 50K domain to move as a separate unconstrained entity (for the

definition of the upper 50K domain, see Supplementary Information). We then achieved a very good fit (Figs 3a and 4a, bottom). Analysis of the movement of the upper 50K domain showed that it was a rotation of 21° about an axis passing close to the α -carbons of amino acids 449 and 468 (chicken rigor sequence; Protein Data Bank accession code 2Mys). Glu 468 is part of switch 2 (for coordinates of the new structure and the rotation axis, see Supplementary Information).

Thus, the cleft in the myosin 50K domain does seem to close on strong binding to actin as previously suggested³. In addition, switch 1 moves away from the nucleotide-binding site on strong binding to actin. Figure 4a shows details of the density near the nucleotide-binding site. The density dips down towards the nucleotide-binding site in a way that is not compatible with the initially fitted crystal structure of the myosin cross-bridge (Fig. 3a, top), whereas the structure with a displaced upper 50K domain lies completely within the density (Fig. 3a, bottom).

The myosin cross-bridge can bind to actin in one of two ways, either weakly or strongly. The initial weak binding to actin isomerizes to strong binding¹⁰. We equate the two structures that we have described with these two states. The lower 50K domain seems initially to make a weak stereospecific interaction (area of contact, $\sim 1,000 \text{ \AA}^2$) with the actin filament involving the lower 50K domain (actin-binding cleft open). The upper 50K domain initially makes no contact with actin. On strong binding, such as occurs in the actin myosin rigor complex, the upper 50K domain seems to swing round so that the cardiomyopathy loop (containing the first reported familial cardiomyopathy missense locus, Arg 405 $\rightarrow$ Gln; ref. 11) comes into contact with the actin surface (total area of contact, $\sim 2,000 \text{ \AA}^2$; Fig. 3). This mutation has been shown to affect the rate of actin-activated ATPase^{12,13}, as is to be expected if this loop is part of the actin-binding site. The switch 1 element of the nucleotide-binding site is anchored in the upper 50K domain, however, so that the strong binding to actin will lead to a 6–9 Å movement of switch 1. Strong binding to actin therefore seems to open the nucleotide-binding pocket (Fig. 4a and b, bottom). This movement is distinct from the opening of switch 2 during the power stroke.

The least-squares analysis (Supplementary Information) yielded the standard deviations for all of the parameters (shift vectors and rotation angles) varied during fitting. From their values (Supplementary Information), we derived a standard deviation for the positioning of the backbone of the polypeptide in the model structure of about 1–2 Å, depending on the distance from the centre of gravity of the upper 50K domain. In the nucleotide-binding pocket, this error is roughly 1.5 Å with an observed movement of 6–9 Å. The resulting molecular models show typical protein interfaces with normal interface distances (for the coordinates, see Supplementary Information).

It has been suggested that the major cleft in the 50K domain will probably close on strong binding³. Electron microscopy studies of smooth muscle myosin bound to actin suggest that the cleft closure is sensitive to ADP binding¹⁴. Two spectroscopy studies show that the actin-binding cleft probably closes on strong binding of the cross-bridge to actin^{15,16}. A crystal structure of a myosin-V head¹⁷ shows that the actin-binding cleft can close even in the absence of actin. In addition, a *Dictyostelium* myosin II crystal structure without nucleotide¹⁸, although not showing the relative disposition of the upper and lower 50K domains that we find here, does show a bending of the β -sheet of the motor domain. This leads to a destruction of the nucleotide-binding pocket. The bending may account for the hinge function that we have postulated.

Our findings suggest that the closing of the actin-binding cleft is mechanically linked to the opening of switch 1. It has been already established that the pre- and post-power stroke states are linked to the opening and closing of switch 2. Combining the opening and closing of switch 1 with the opening and closing of switch 2 results in four

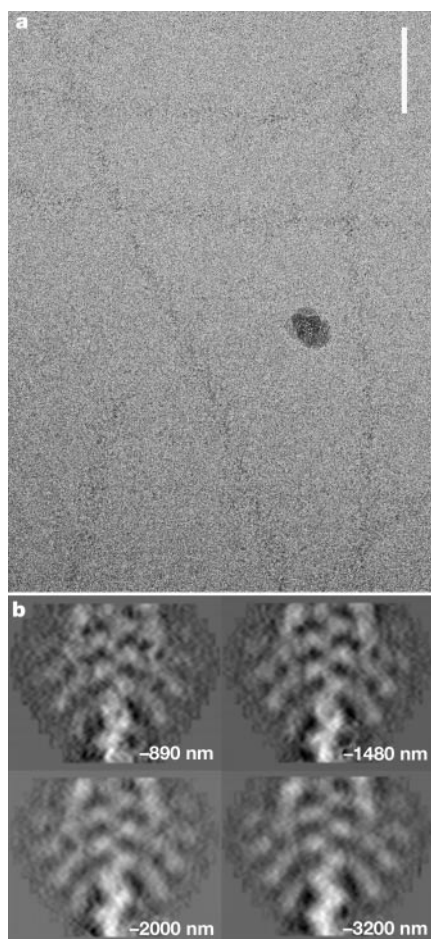


Figure 1 Electron cryo-microscopy and image processing of the complex of F-actin and myosin subfragment 1 (decorated actin). All images were recorded by zero-loss energy-filtered cryo-microscopy (Methods). Increased amplitude contrast of native protein allows imaging at lower dose and very close to focus. This is used to record defocus series of up to four exposures, which are then combined to correct for imaging aberrations (correction of contrast transfer function; Methods). **a**, Typical sample of decorated actin at an underfocus of $1.48 \mu\text{m}$ (the scale bar of 77 nm equals one helical repeat of the 28/13 helix). **b**, Averages of 104 identical object areas selected from four exposures in one defocus series. Half a helical repeat (38.5 nm) is shown for each underfocus average.

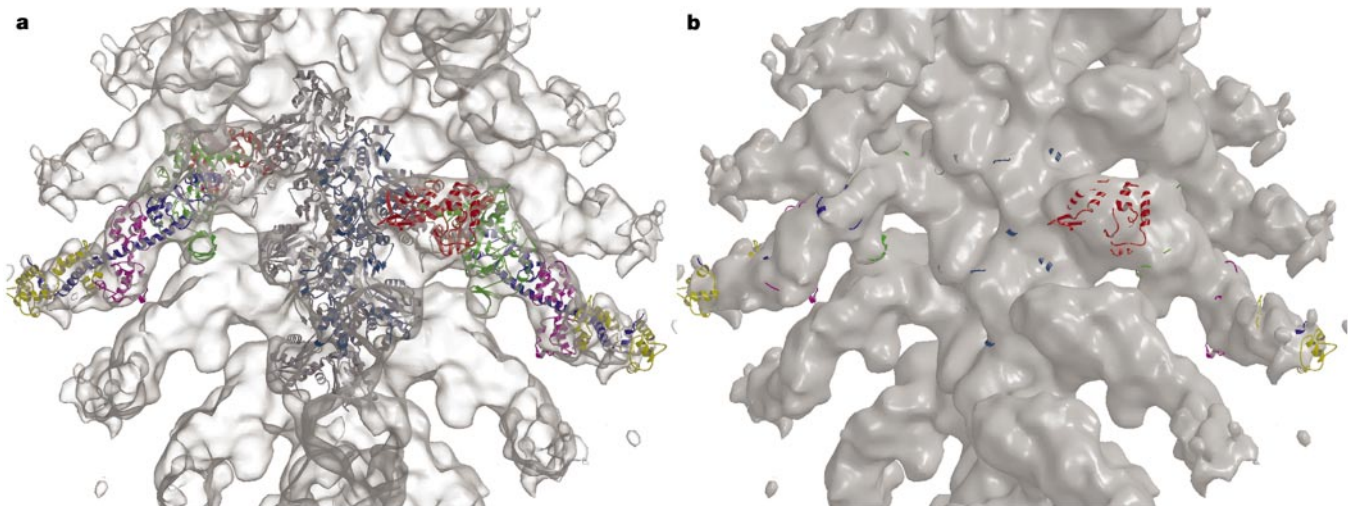


Figure 2 Fit of crystallographic molecular models of F-actin and myosin subfragment 1 into the reconstructed density. The molecular models (ribbon representation) of myosin and F-actin are shown docked into the experimental density (electron cryo-microscopy reconstruction shown at a cut-off level of 2σ). The fit was achieved by using an iterative least-squares algorithm (Methods). **a**, General fit of F-actin and myosin (two cross-bridges

are shown). Here, myosin has been kept in its original X-ray crystallographic conformation without bound nucleotide. The density contoured at one-third of the maximum is transparent. **b**, As **a**, but with a solid surface. The surface representation shows the localized poor fit of the myosin upper 50K domain (shown in red), which protrudes out of the density.

states, which can be mapped onto those of the Lymn–Taylor scheme¹ to produce a four-state model of the cross-bridge cycle (Fig. 5). In the pre-power stroke conformation the switch 2 element forms hydrogen bonds to the γ -phosphate, and in the post-power stroke it does not (for example, see ref. 2). By contrast, we now see a movement of the switch 1 rotating away from the phosphate-binding site.

In the related motor protein kinesin¹⁹ the situation is the other way round: although the strong coordination of the phosphate

groups by switch 1 would seem to be a necessary adjunct to hydrolysis, in all published crystal structures the switch 1 element is essentially ‘open’ and the strongly coordinated form is missing. These facts fit well with the different functions of myosin and kinesin. Whereas the binding of the myosin cross-bridge to actin causes the release of products, the binding of kinesin to the microtubule enables hydrolysis. In fact, there is now evidence from electron paramagnetic resonance spectroscopy that switch 1

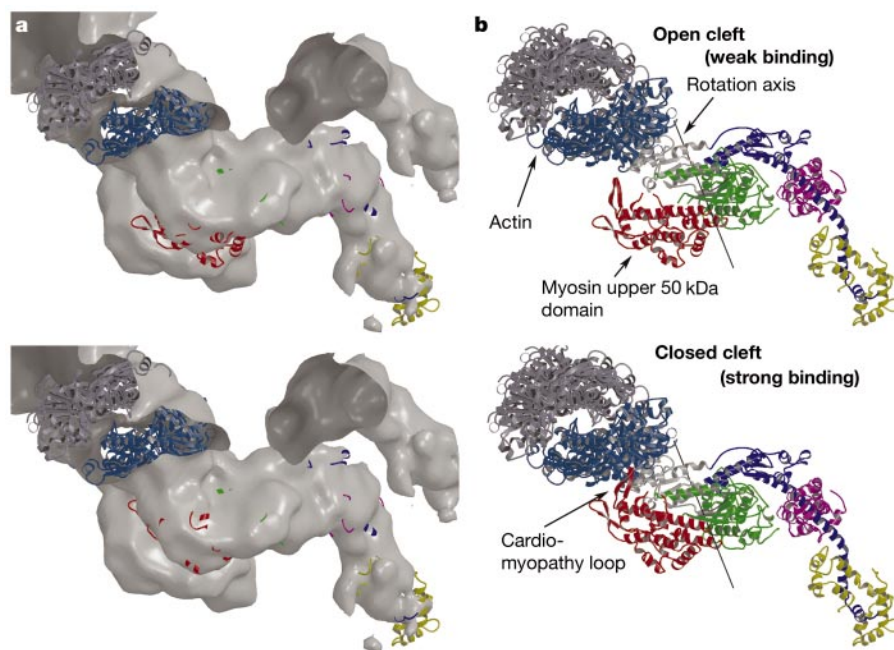


Figure 3 Conformational change in myosin on strong binding to F-actin. The conformational changes necessary to obtain a perfect docking of the molecular model of myosin S1 into the reconstructed density are shown here. **a**, Two end-on views of the F-actin axis at the same cut level as in Fig. 2, but with two fits of the myosin ribbon model to the density (dark grey surface represents the inner skin of the surface model, which is visible because the surface rendering is cut open, as it is for the adjacent myosin along the

helix). Top, fit with the original X-ray structure, showing the upper 50K domain (red) clearly outside the density. Bottom, fit with an independent least-squares docking of the upper 50K domain. **b**, Overview of the docking obtained with a rigid-body X-ray model of myosin (top) and a fit with an independent upper 50K domain (bottom). The area in contact with actin increases from $1,000 \text{ \AA}^2$ to $2,000 \text{ \AA}^2$.

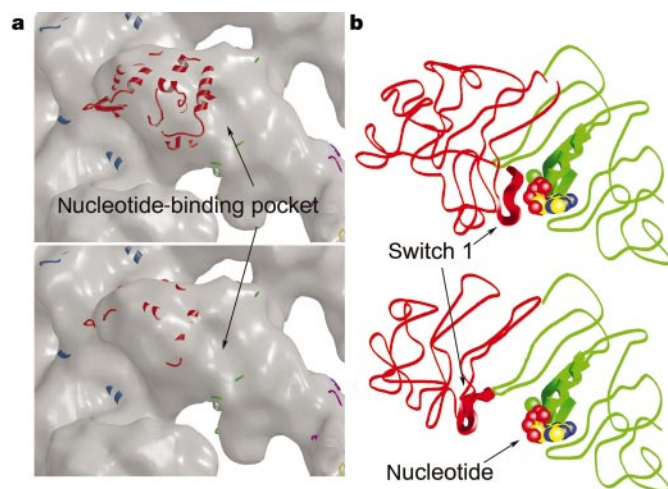


Figure 4 Movement of the nucleotide-binding pocket switch 1 element in myosin. Detailed views are presented at right angles to the actin filament axis, showing the conformational changes resulting from fitting the upper 50K domain into the experimental density. **a**, Top, global fit of the putative weak binding state of myosin to actin, with a large portion of the upper 50K domain outside the experimental density. Arrows indicate the nucleotide-binding site, which lies in a hollow in the density. Bottom, allowing a movement of the upper 50K domain leads to a perfect fit into the density. **b**, Changes of the molecular structure in the nucleotide-binding site arising from the movement shown in **a**. Switch 1 is indicated.

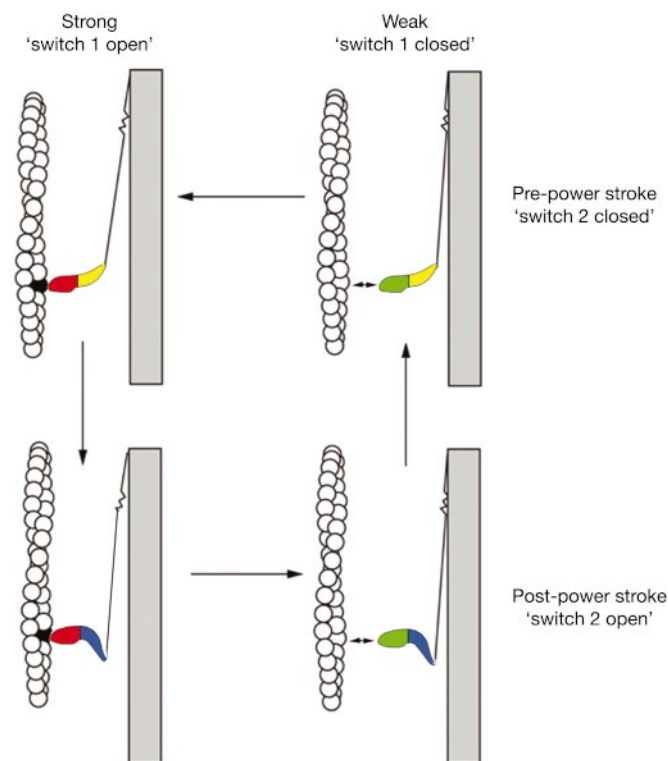


Figure 5 Switch 1 in-out and switch 2 in-out movements mapped onto the Lymn-Taylor cross-bridge cycle. The in-out movement of switch 1 is coupled to weak-strong actin binding. The in-out movement of switch 2 is coupled to movement of the lever arm (for example, see ref. 2). Combining these possibilities gives rise to four states that may be equated with the states predicted in the Lymn-Taylor cross-bridge cycle¹.

closes when kinesin binds to microtubules²⁰. For myosin it would seem advantageous for switch 1 to be closed until opened by the binding to actin; thus, the myosin cross-bridge alone would be naturally found in the 'switch 1 closed' form. For kinesin, by contrast, the binding to the microtubule may enable hydrolysis by bringing switch 1 into contact with the polyphosphate moiety of ATP; thus, it would be advantageous for kinesin alone to be in the 'switch 1 open' form. □

Methods

Preparation of actin and myosin subfragment 1

Chicken myosin S1 was purified as described²¹ and stored as a frozen solution (25 mg ml⁻¹) in 30% sucrose (w/v) at -80 °C. The final concentration of S1 in the microscope sample was about 0.5 mg ml⁻¹. Actin from rabbit muscle²² was mixed with myosin S1 in Eppendorf tubes at a ratio of 1:3 (w/w) in decoration buffer (50 mM KCl, 2 mM MgCl₂, 20 mM Tris-HCl (pH 7.3), all final concentrations). It is well known that the RLC can be lost under these treatments. We did not try to quantify the ratio of the RLC to the heavy chain, but it is clear from difference maps of the reconstructed density and the molecular model that about 50% of the RLCs are missing (Supplementary Information). The suspension of decorated actin filaments was directly used for preparing samples for electron cryo-microscopy.

Making grids for cryo-microscopy

Samples were prepared as thin layers over holes and vitrified by rapid freezing in a mixture of liquid ethane and liquid propane. Carbon film with holes was prepared as described²³. Sample layers were produced by pipetting, blotting and freezing in a computer-controlled environmental chamber that maintained 100% humidity at room temperature²⁴.

Electron cryo-microscopy

Samples were studied using an energy-filtering TEM LEO 912 electron microscope (Omega) at 120-kV electron energy fitted with a lanthanum boride source. Two 3500Z cryo-transfer sample holders (Oxford) were used, one with the conventional sample tip and one modified with an exchangeable sample cartridge. All data used for this study were recorded on imaging plates (Fuji), which were then digitized in a Micron 1 imaging plate scanner (Ditabis) at a pixel size of 25 µm. Elastic bright field images of decorated filaments were recorded in zero-loss mode⁷. By using a model 974 ssCCD (Gatan) for low-dose focusing, we were able to collect defocus series of individual sample areas consisting of four exposures. The nominal underfocus values used for imaging were 0.5, 1.0, 1.5 and 2.5 µm, accumulating a total electron dose of 2,000 electrons per nm² of the sample area (four exposures at 500 electrons per nm²). Inspection of the defocus series showed a relatively small variation of the defocus values; the data used in this study cover an underfocus range from 430 to 3,200 nm. The reconstruction presented here is based on 15 selected defocus series.

Single helical repeat image processing

Images of decorated filaments were analysed and processed using the software package SPIDER²⁵ and a suite of programs specifically designed for single particle analysis and reconstruction of helical filaments. The algorithms and processing schemes used are developments of the conventional single helical repeat image processing²⁶, the multivariate statistical analysis used for projection angle determination²⁷, and a helical reconstruction algorithm in real space by diagonalized least squares with an axial tomographic description²⁸. We used a mask to restrict the volume of the reconstruction. A general outline of the processing steps is given below (for additional information, see Supplementary Information).

The reconstruction is derived from 15 defocus series with a total of 1,750 helical repeats, each consisting of 28 actin-myosin dimers forming a helix with 28/13 symmetry and a helical repeat of 2 × 38.5 nm (total of about 49,000 'single' actin-myosin complexes). Only these helical repeats were included in the final image processing and 3D reconstruction, which showed identical helical classification and Euler angle assignment in all four exposures of one defocus series (for details, see Supplementary Information).

Figure preparation

The programs BobScript v2.6 (ref. 29), Molscrip (ref. 30) and Raster 3D (ref. 31) were used to prepare the figures.

Received 11 March; accepted 21 August 2003; doi:10.1038/nature02005.

1. Lymn, R. W. & Taylor, E. W. Mechanism of adenosine triphosphate hydrolysis by actomyosin. *Biochemistry* **10**, 4617-4624 (1971).
2. Geeves, M. A. & Holmes, K. C. Structural mechanism of muscle contraction. *Annu. Rev. Biochem.* **68**, 687-727 (1999).
3. Rayment, I. *et al.* Structure of the actin-myosin complex and its implications for muscle contraction. *Science* **261**, 58-65 (1993).
4. Schröder, R. R. *et al.* Three-dimensional atomic model of F-actin decorated with *Dictyostelium* myosin S1. *Nature* **364**, 171-174 (1993).
5. Volkman, N. *et al.* Myosin isoforms show unique conformations in the actin-bound state. *Proc. Natl. Acad. Sci. USA* **100**, 3227-3232 (2003).
6. Rayment, I. *et al.* Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science* **261**, 50-58 (1993).
7. Smith, C. A. & Rayment, I. X-ray structure of the magnesium(II)-ADP-vanadate complex of the

- Dictyostelium*–*Discoideum* myosin motor domain to 1.9 Å resolution. *Biochemistry* **35**, 5404–5417 (1996).
8. Dominguez, R., Freyzon, Y., Trybus, K. M. & Cohen, C. Crystal structure of a vertebrate smooth muscle myosin motor domain and its complex with the essential light chain: visualization of the pre-power stroke state. *Cell* **94**, 559–571 (1998).
 9. Angert, I., Majorovits, E. & Schröder, R. R. Zero-loss image formation and modified contrast transfer theory in EFTEM. *Ultramicroscopy* **81**, 203–222 (2000).
 10. Geeves, M. A. & Conibear, P. B. The role of three-state docking of myosin S1 with actin in force generation. *Biophys. J.* **68**, 194S–201S (1995).
 11. Geisterfer-Lowrance, A. A. *et al.* A molecular basis for familial hypertrophic cardiomyopathy: a β cardiac myosin heavy chain gene missense mutation. *Cell* **62**, 999–1006 (1990).
 12. Sweeney, H. L., Straceski, A. J., Leinwand, L. A., Tikunov, B. A. & Faust, L. Heterologous expression of a cardiomyopathic myosin that is defective in its actin interaction. *J. Biol. Chem.* **269**, 1603–1605 (1994).
 13. Sasaki, N., Asukagawa, H., Yasuda, R., Hiratsuka, T. & Sutoh, K. Deletion of the myopathy loop of *Dictyostelium* myosin II and its impact on motor functions. *J. Biol. Chem.* **274**, 37840–37844 (1999).
 14. Volkmann, N. *et al.* Evidence for cleft closure in actomyosin upon ADP release. *Nature Struct. Biol.* **7**, 1147–1155 (2000).
 15. Yengo, C. M., De la Cruz, E. M., Chrin, L. R., Gaffney, D. P. & Berger, C. L. Actin-induced closure of the actin-binding cleft of smooth muscle myosin. *J. Biol. Chem.* **277**, 24114–24119 (2002).
 16. Conibear, P. B., Bagshaw, C. R., Fajer, P., Kovacs, M. & Malnasi-Csizmadia, A. Myosin cleft movement and its coupling to actomyosin dissociation. *Nature Struct. Biol.* (in the press).
 17. Coureux, P.-D. *et al.* A structural state of the myosin V motor without bound nucleotide. *Nature* **425**, 419–423 (2003).
 18. Reubold, T., Eschenburg, S., Becker, A., Kull, F. J. & Manstein, D. J. A structural model for actin-induced nucleotide release in myosin. *Nature Struct. Biol.* (in the press).
 19. Kull, F. J., Vale, R. D. & Fletterick, R. J. The case for a common ancestor: kinesin and myosin motor proteins and G proteins. *J. Muscle Res. Cell Motil.* **19**, 877–886 (1998).
 20. Naber, N. *et al.* Closing of the nucleotide binding pocket of kinesin-family motors upon binding to microtubules. *Science* **300**, 798–801 (2003).
 21. Margossian, S. S. & Lowey, S. Preparation of myosin and its subfragments from rabbit skeletal muscle. *Methods Enzymol.* **85**, 55–71 (1982).
 22. Spudich, J. A. & Watt, S. The regulation of rabbit skeletal muscle contraction. I. Biochemical studies of the interaction of the tropomyosin–troponin complex with actin and the proteolytic fragments of myosin. *J. Biol. Chem.* **246**, 4866–4871 (1971).
 23. Jahn, W. Easily prepared holey films for use in cryo-electron microscopy. *J. Microsc.* **179**, 333–334 (1995).
 24. Löbau, V. *Neue Wege in der Kryo-Präparation: Die Controller-gesteuerte Einfrierapparatur*. Thesis, Univ. Stuttgart (1997).
 25. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
 26. Egelmann, E. H. A robust algorithm for the reconstruction of helical filaments using single particle methods. *Ultramicroscopy* **85**, 225–234 (2000).
 27. Schroeder, R. R., Angert, I., Frank, J. & Holmes, K. C. Multivariate statistical analysis and tomographic processing of helical objects. *Proc. 15th Int. Congr. Elec. Microsc.* **3**, 443–444 (2002).
 28. Yagle, A. E. in *Discrete Tomography* (eds Herman, G. T. & Kuba, A.) (Birkhäuser, Boston, 1999).
 29. Esnouf, R. M. Further additions to Molscrip version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55**, 938–940 (1999).
 30. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
 31. Merritt, E. & Bacon, D. Raster 3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Frank for discussions, support and encouragement, and P. Bele for technical assistance during image processing.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.C.H. (holmes@mpimf-heidelberg.mpg.de).

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building

4 Crinan Street

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Global Head of Naturejobs:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Spain/

Portugal/ Belgium:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Natureevents:

Paul Constant (4954)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejpn.com

naturejobs

True grit

My decision to finish my undergraduate education at the University of Sheffield, UK, in 1991 was met with scepticism from my friends. But I handled it well, I thought, saying that Sheffield couldn't be any grungier than Milwaukee, the blue-collar city where I had started my degree before the opportunity to study abroad arose. So I packed my bags and traded one industrial city that had seen better days for another — steeling myself for the inevitable *Laverne and Shirley* references once I arrived in Yorkshire.

Although I had prepared myself for the worst, I experienced the best. The people were friendly and the countryside bleakly beautiful. And one memorable weekend of hitchhiking with a friend took me to Manchester and Liverpool, both of which I thought had a gritty charm — just like Milwaukee and Sheffield.

When I decided this year to revisit northern England, I wasn't sure how much things would have changed. Some of my work colleagues had the same preconceptions as my friends had in 1991. But returning to Manchester, Liverpool and Sheffield was like experiencing the moment in *The Wizard of Oz* when the film switches from black-and-white to colour. The changes went deeper than the removal of soot and the conversion of derelict warehouses into expensive apartments. It was evident in the spirit of investment in the scientific infrastructure and the willingness between the cities to collaborate rather than compete (see overleaf).

The experience, I think, extends beyond northern England. Attitudes of reverse provincialism (a polite phrase for snobbishness) are rife in the Western world. Scientific centres such as Oxford, Cambridge, Boston and San Francisco are justifiably worthy of note. But there are many other lower-profile areas where sound science is done and good jobs are available. Grit, rather than just reputation, can transform anywhere into a good place to do science.

Paul Smaglik

Naturejobs editor



Contents

REGIONS

Rivitalizing the north
of England

p430

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Rising star Northern England



Angel of the North: this colossal statue greets travellers on the road to Newcastle — and perhaps reaches out to the biotech industry's very own 'business angels'.

Northern England was the crucible for an industrial revolution around 200 years ago — although it is a revolution upon which the sun has now set. But in the shadows of defunct textile, coal and steel industries, a new revolution is being forged. At its hub is a cluster of universities, businesses and economic development agencies, drawing in an unprecedented amount of infrastructure investment. The area is vibrant with biomanufacturing, new academic buildings, industrial facilities and incubators. As alliances are forged and investment flows into the region, the north is beginning to build its own version of the southeast's 'golden triangle'.

The southern triangle stretches between Cambridge, Oxford and London and has both history and a certain cachet on its side. For the north, the area is roughly defined by Liverpool, Sheffield and Newcastle upon Tyne — and is bolstered by a combative determination combined with a willingness to work together (see 'Joint efforts', page 432).

But buildings alone won't transform the region. It needs young companies to fill its incubators, in a

period when venture capital is proving unadventurous. And it needs to overcome a general perception that it is an underdog in order to attract top talent to its universities.

NORTH BY NORTHWEST

Travelling the perimeter of the triangle and visiting the institutions along the way reveals that the region is making good progress towards its goals. Manchester makes a good starting point. It is the heart of the current revolution (even if not at the geographical centre) and is in many ways its driving force. Key to the city's scientific rejuvenation is the impending merger of its two universities — the University of Manchester and the University of Manchester Institute of Science and Technology (UMIST) — which is scheduled for October next year (see 'Teething troubles for transferred property rights', below).

Architects of the merger say that they want the combined institution to be greater than the sum of its parts. That expansion is evident in the business plan, which calls for over 100 new appointments, says

Teething troubles for transferred property rights

The succession of power at the soon-to-be merged University of Manchester and the University of Manchester Institute of Science and Technology (UMIST) was easily arranged. Both John Garside, UMIST's principal and vice-chancellor, and his opposite number at the University of Manchester, Martin Harris, agreed to retire within a year of the arrival of Adam Gilbert, who will leave the top slot at the

University of Melbourne, Australia, to run the new entity next February. But another transfer — of two different intellectual property (IP) portfolios and policies into a single unit — may prove to be more problematic and have major bearings on how the new university will create new companies and jobs.

At UMIST, IP is owned by the academics. At the University of Manchester, the

university owns it, but the academics who generate it and the administrators who manage it tend to get a cut. The Northwest Development Agency is trying to steer the new IP office towards the University of Manchester model. The agency will fund the new office with a £30-million (US\$48-million) grant — but may withhold the money if it isn't satisfied with the choice of management or business model.

P.S.

Michael Grant, pro-vice-chancellor of the University of Manchester. Some of these new positions are likely to be in areas identified for growth by the United Kingdom's Office of Science and Technology, including photon science and neuroscience. Others will come out of a concerted effort to combine UMIST's engineering expertise with the University of Manchester's strength in basic sciences. The merger is creating a chance to "break out of universities that are solely organized around traditional disciplines", says John Garside, principal and vice-chancellor of UMIST.

Perhaps the most tangible example of such thinking is one of the university-to-be's most ambitious building projects, the Manchester Interdisciplinary Biocentre (MIB). Proposed by UMIST in 1998, before merger talks began, the £35-million (US\$57-million), 13,700-square-metre building will house more than 500 scientists in some 85 research groups. They will be working in one of six themes: systems biology; biomolecular and cellular analysis; biomolecular structure and dynamics; bionanotechnology; biophotonics and bioelectronics; and biomaths and biocomputation. "This centre will be home to scientists and engineers from across the disciplines," says John McCarthy, leader of the MIB project. The aim is to make biology more quantitative by bringing chemists, engineers, physicists and mathematicians into the mix.

The MIB is the first of possibly three or more multidisciplinary science buildings for the new university. Proposals for a neuroscience building and photonics facility are advancing through the pipeline.

Other projects already under way are concentrated in a life-sciences hub centred on the current University of Manchester campus. The £18-million Integrated Centre for Molecular and Cellular Biology is due to open in January 2004, although the university is still looking for funds to start construction on its fourth wing. The university's incubator, full after only two years, is to double in size. And in mid-2004, the £22-million Wolfson Molecular Imaging Centre will open at Christie Hospital.

All these initiatives mean that Manchester has changed a lot in ten years, says Mark Ferguson, chief executive of Renovo, a biotech company specializing in wound-healing. When Ferguson was dean of the University of Manchester's biosciences department, he helped to establish the incubator in which Renovo is

now housed. He can prove that being based in the north (which until a few years ago was considered a disadvantage) need not hinder a biotech company from recruiting talent and raising money. Last year, the company raised £23 million in private venture funds in its second round of financing. This summer Ferguson recruited Andrew Kay, former worldwide head of marketing for drug firm Novartis, from Basel in Switzerland.

The building and recruitment boom in the Manchester area is not limited to biotech companies or universities. A cancer-research facility being built at drug company AstraZeneca's Macclesfield site will house 200 scientists, primarily chemists and biologists, when it opens in 2005. The firm has recently invested some \$150 million in the area for upgrades, including more high-throughput screening for potential drug compounds and an automated compound library.

That investment, and the positions it will create, echoes what is happening throughout the northwest. "There are going to be significant job opportunities here for scientists," says John Stageman, AstraZeneca's vice-president for enabling science and

technology. Stageman is also a member of the North West Science Council, the United Kingdom's first regional science council.

One of the northwest's biggest recent coups is securing £30 million in public and private funds in March to build a biomanufacturing centre in Speke, near Liverpool, which will be run by Eden Biodesign, a consultancy that specializes in biomanufacturing. The plant is intended

to employ about 100 people to manufacture biologics, such as proteins, vaccines and antibodies, in small amounts for start-up companies. Its location, just down the road from drug companies Chiron, MedImmune and Eli Lilly, will increase the area's reputation as a hotbed of biomanufacturing employment.

The facility's capacity to manufacture vaccines should complement the region's strengths in immunology, which are also on the rise. The long-established Liverpool School of Tropical Medicine has been expanding since it helped to analyse the genomes of the malaria parasite and its mosquito host last year.

"We're basically building on the back of that," says director Janet Hemingway. The school's staff, 168-strong in 2001, is on course to double by 2006, with increased opportunities for students and postdocs. The expansion is being funded with large-scale grants by



Expanding roles: biotech research at the Liverpool School of Tropical Medicine (right) and at Manchester (above) is helping northern England to build on its scientific strengths.





The Centre for Developmental Genetics at the University of Sheffield has expanded in the past few years.

the UK government, the Wellcome Trust, the Bill and Melinda Gates Foundation in Seattle, Washington, and the US National Institutes of Health. The school is also raising funds for a £24-million addition that would improve its proteomics capabilities.

Liverpool has a history of biology — pro-vice-chancellor Julian Crampton, professor of molecular biology, says that the University of Liverpool devotes about 45% of its resources to biomedical research. But only within the past three years has the university started thinking about spinning out and courting companies. One result of this exercise, the Merseyside Biotechnology Incubator, opened in April and already has three companies inside, with room for 12 more.

The university and its neighbour, Liverpool John Moores University, are also planning a science park, which will have a broader tenant base than the bioincubator. Its headquarters will sit in the shadows of the famous spaceship-shaped Liverpool Metropolitan Cathedral.

Reaching Newcastle, the tip of the northern

triangle, reveals another set of recent and ongoing developments. The £19-million Institute for Research in Environment and Sustainability, which will open at the University of Newcastle upon Tyne late this year, has room for 500 staff. Many will be moved in from other university buildings, but new multidisciplinary positions in environmental genomics and proteomics will probably be filled from outside. There will also be room for social scientists to interact with the environmental scientists. There is no point, says the institute's director Tony O'Donnell, in designing environmentally friendly technologies such as hydrogen-powered cars if the technology is cumbersome or expensive.

That project is joined by continued growth to another young one, the Centre for Life, a facility in downtown Newcastle that, from a distance, resembles a football stadium — until you see the giant metallic spider crawling up its side. The centre combines a children's science museum — with lab benches in different rooms for different heights and ages — with genetics research and a fertility clinic. The grouping is not accidental. The fertility clinic provides embryos for stem-cell research, which can then be studied by the genetics groups — while children can watch the lab scientists at work.

Alison Murdoch, director of the Newcastle Fertility Centre within the Centre for Life, says that the set-up has improved stem-cell science in the city. "Since we've moved, we've developed closer links to the geneticists," she says. That link is important, because the centre is one of the few in the United Kingdom allowed to

Joint efforts

Although many of the developments in the north are centred on single institutions, some of the more ambitious schemes involve multiple players — a strategy necessary to keep up with the south.

One of the largest is Manchester's National Genetics Reference Library, funded by a £60-million (US\$97-million) grant from the health department for the University of Manchester and Christie Hospital to generate a genetic history of all patients over the age of 45 by monitoring them until their deaths. The programme will be run by the local health-service trust and Mark Ferguson, former dean of the University of Manchester's biology department. Getting the grant showed that the north can be competitive against the south.

Meanwhile, several northwestern entities are grouping basic genetics, clinical science and public knowledge under the umbrella of the Northwest Genetic Knowledge Park. They have already raised about £100 million in public funds for a Manchester building, and anticipate the entire project costing about £280 million.

In a different collaboration, David Williams, a clinical-engineering professor

at the University of Liverpool, has set up a tissue-engineering programme between Liverpool and Manchester. Combining the engineering tradition of the University of Manchester Institute of Science and Technology (UMIST) with the University of Manchester's basic-science background and Williams' leadership in the field makes tissue engineering a regional strength, says Julian Crampton, pro-vice-chancellor of the University of Liverpool.

Sheffield, Leeds, UMIST and the University of Manchester have also joined forces to form the North of England Structural Biology Centre to collaborate on nuclear magnetic resonance spectroscopy and X-ray crystallography, two expensive technologies used to solve protein structure. The group is also pushing to build a new synchrotron light source, known as the Fourth Generation Light Source, at Daresbury, as the existing one there is being supplanted by Diamond in the south which goes online in 2006.

The Liverpool School of Tropical Medicine has a history of reaching beyond its walls — a trait that is accelerating now that its funding has increased. The school's staff looks after the infectious-disease ward

at the Royal Liverpool University Hospital and Alder Hey Hospital. The school is also stepping up its private-sector collaborations, spinning out one company and developing and launching a new antimalarial drug with GlaxoSmithKline.

Under the Consortium for Post-Genome Science, the University of Manchester, the University of Liverpool, UMIST, the University of Salford and the Daresbury Laboratory are pooling their resources and competing for grants and sharing expensive equipment such as microarrays, says Andrew Cossins of the Laboratory for Environmental Gene Regulation at the University of Liverpool, and a member of the consortium's bioarray programme. The group's philosophy extends to the entire north. "The region will sink or swim on its ability to pull together rather than relying on single institutions," Cossins says.

National Genetics Reference Libraries

♦ www.ngri.co.uk

UK Centre for Tissue Engineering

♦ www.ukcte.org

The Consortium for Post-Genome Science

♦ www.postgenomeconsortium.com

North West Genetic Knowledge Park

♦ www.nowgen.org.uk

P.S.

Web of wonder: the Centre for Life in Newcastle is building strong bonds between genetics and fertility research.



derive embryonic stem cells for research. This group is growing rapidly thanks to the interest it has aroused among students and postdocs, Murdoch says. And John Burn, executive director of the Northern Genetics Knowledge Park at the centre, also expects growth for his team of scientists there — particularly in the research area of genetic instability.

WHITE ROSES

Heading south and east from Newcastle into Yorkshire brings you into the territory of the White Rose University Consortium. The universities of Leeds, York and Sheffield have formed a loose alliance — named after the county's historic emblem — and hope to foster developments equal to those of their counterparts across the Pennines in Manchester and Liverpool. This is where scientists are most candid about their prospects of competing with the south.

Leeds, like many other universities in the north, is focusing on interdisciplinary research and investing accordingly. Construction began this August on a £12-million proteomics centre, strategically placed between the medical school and the biological sciences building. The university is also spending £4.5 million on a bionanotechnology centre.

John Findlay, a biochemistry professor at the University of Leeds, sees this project as a positive start, but argues that in order to attract industry and jobs, the area will have to do more. "The talent is here," Findlay says. "It's just that the traditional investment hasn't been here."

The money is slowly beginning to trickle in. York's science park, open for only six months, is filling up with small companies. One of these, Xceleron, which uses its own accelerator and mass spectrometer to help big drug companies to characterize 'lead' compounds and targets, is now looking for more nuclear physicists to assist in running its machinery. Colin Garner, the company's chief executive, is also looking for a clearer sense of leadership for the greater-northern region. A group of at least half a dozen organizations is trying to pull the area together, but he thinks that more centralization will be necessary.

Some drive could come from the few larger companies located in Yorkshire. York-based medical-device company Smith & Nephew is providing some

by funding studentships at the University of York and fostering research collaborations there.

Peter Arnold, director of technology at Smith & Nephew, says that many people have preconceptions about the medical-device company, just as they do about the state of science in northern England. Medical-device research is now much more high-tech than it once was, he says, as it now focuses more on tissue engineering and materials science than on clunky plastic orthotics; 120 of the site's 180 employees hold PhDs.

In nearby Harrogate, another large company is also hungry for freshly minted PhDs. Within the past year Covance, the global contract research organization, opened a new animal-research facility there, and the firm has since been searching for scientists interested in conducting preclinical research.

In Sheffield, meanwhile, the building work is just beginning. The European Regional Development Fund is pouring £11 million over three years to develop biology. The South Yorkshire Bioscience Enterprise Network is planning an incubator, and a steady stream of spin-offs from both the University of Sheffield and Sheffield Hallam University could help to fill one.

Despite all of the progress, some of the old prejudices remain. For example, some venture capitalists funding a spin-off from the University of Sheffield urged the company's management to move its base south.

Phil Ingham, head of the biomedical science department and director of the Centre of Developmental Genetics at the University of Sheffield, says that much has changed since he was recruited in 1996 from the Imperial Cancer Research Fund (now part of Cancer Research UK) in London. He was able to raise £30 million to gut, restore and transform one of the university's oldest buildings into one of its most state-of-the-art. He has also expanded the staff from eight principal investigators when he started to 12 now, with more growth in sight. Even so, he says, it will take further effort to gain the respect that more prestigious institutions in the golden triangle have. But he sees progress, nonetheless.

"We're going places," Ingham says. Anyone who has driven around the triangle will have seen that the whole region is already on its way.

Paul Smaglik is editor of *Naturejobs*.